

1. 33 No. 1

(New Series No. 151)

October 1972

---

# ANALYSIS

---

Cue - H0 2966 - 3 - P00 6339

108  
An 13

Edited by

C. J. F. WILLIAMS

---



## CONTENTS.

You and I	C. L. HAMBLIN
The prescriptivist definition of 'better'	M. B. SMYTH
Inverse discrimination	J. L. COWAN
Reasons for being moral	RODGER BEEHLER
On giving reasons for being moral	KAI NIELSEN
Morals and reasons	RODGER BEEHLER
<sup>Reference</sup> 'Bedeutung' in Frege: a reply ✓	V. H. DUDMAN
On seeing everything upside down	N. G. E. HARRIS
Brady on recklessness	WILLIAM J. WINSLADE

---

BASIL BLACKWELL · ALFRED STREET · OXFORD

---

*Forty pence*

# ANALYSIS

Edited by C. J. F. Williams



3

Volume 33  
1972-1973

BASIL BLACKWELL · ALFRED STREET · OXFORD

# INDEX

	PAGE
ALTMANN, D. R.	Geach on 'know' in 'if' clauses 174
ANSCOMBE, G. E. M. <i>and</i>	
FELDMAN, J.	On the nature of justice in a trial 33
ATTFIELD, ROBIN	How things exist: a difficulty 141
BAYLY, MICHAEL D.	Reparations to wronged groups 182
BEEHLER, RODGER	Reasons for being moral 12
BEEHLER, RODGER	Morals and reasons 19
BERGSTRÖM, LARS	On the coherence of act-utilitarianism 98
BOËR, STEVEN E.	Russell on classes as logical fictions 206
CAMPBELL, T. D.	Formal justice and rule-change 113
CARTER, W. R.	On promising the unwanted 88
CHERRY, CHRISTOPHER	Two views of moral practices 118
COWAN, J. L.	Inverse discrimination 10
DAUER, FRANCIS W.	How not to reidentify the Parthenon 63
DAVIS, LAWRENCE H.	Smart on conditions of identity 109
DUDMAN, V. H.	'Bedeutung' in Frege: a reply 21
DUMMETT, MICHAEL	Frege's way out: a footnote to a footnote 139
FELDMAN, J.	(See Anscombe, G.E.M.) 33
GEDDES, LEONARD	On the intrinsic wrongness of killing innocent people 93
GERT, BERNARD <i>and</i>	
MARTIN, JAMES A.	105 AM13 'What a man does he can do'? 168
GERT, BERNARD <i>and</i>	
MARTIN, JAMES A.	Outcomes and abilities 188
GOWER, BARRY	Martin on explanation and confirmation 107
GREEN, O. H.	Some supposed advantages of materialism 124
HAMBLIN, C. L.	You and I 1
HARRIS, N. G. E.	On seeing everything upside down 28

P, 6339

HARRISON, ROSS	Lost times	65
ISEMINGER, GARY	Aesthetic judgments and non-aesthetic conditions	129
JOHNSON, CONRAD D.	Davidson on primitive actions that cause deaths	36
JONES, H. M.	Censorship, state secrets and participatory democracy	143
KEARNEY, R. J.	Meaning and implication: other thought	47
KIRK, R.	Underdetermination of theory and indeterminacy of translation	195
LA CROIX, RICHARD R.	The incompatibility of omnipotence and omniscience	176
LEAHY, MICHAEL	Justice in a trial	192
LEWIS, CARROLL	On undetectable differences in sensations	193
LYCAN, WILLIAM G.	Davidson on saying that	138
MARTIN, JAMES A.	(See Gert, Bernard)	168
MARTIN, JAMES A.	(See Gert, Bernard)	188
MARTIN, MICHAEL	Are cognitive processes and structure a myth?	83
MILLER, BARRY	Proper names and <i>suppositio personalis</i>	133
MILLER, DAVID	The truth-likeness of truthlikeness	50
MONRO, HECTOR	Nonsense? Nonsense!	167
NAYLOR, ANDREW	On the evidence of one's "memories"	160
NIELSEN, KAI	On giving reasons for being moral	17
PAGE, EDGAR	Senses of 'obliged'	42
SCHUMM, GEORGE F. and STEVENSON, JOHN G.	Harman's non-essential property	112
SHINER, ROGER A.	Individuals, groups and inverse discrimination	185
SMYTH, M. B.	The prescriptivist definition of 'better'	4
STEVENSON, JOHN G.	(See Schumm, George F.)	112



6761 NON 8

SWARTZ, NORMAN	Is there an Ozma-problem for time?	77
SWINBURNE, R. G.	Confirmability and factual meaningfulness	71
SZABADOS, BÉLA	Wishful thinking and self-deception	201
TAYLOR, PAUL W.	Reverse discrimination and compensatory justice	177
THALBERG, IRVING	Ingredients of perception	145
URMSON, J. O.	Russell's incomplete symbols	111
WALKER, A. D. M.	Goodness of a kind and goodness from a point of view	156
WEDEKING, GARY A.	Reasons for acting versus reasons for believing	102
WINSLADE, WILLIAM J.	Brady on recklessness	31
YOUNG, JOHN J.	Ifs and hooks: a defence of the orthodox view	56

## YOU AND I

By C. L. HAMBLIN

105

AN13

NCL 33 NCL 1-6

OCT. 1972-

JUN. 1973



HOW do I know that I am 'I' and you are 'you'? A typical philosopher's-question, concocted apparently in an attempt to out-doubt Descartes. But the quote-marks around 'I' and 'you' may indicate to the perceptive reader that it is meant as a linguistic question. And, to be precise, I am not unequivocally 'I' at all: I am 'I' only when it is I who speak, not when it is someone else. How (the question should go) do I know that when I say 'I' I mean the same as you do when you say 'you' to me, and that when I say 'you' to you I mean the same as you do when you say 'I'?

One reason this may represent a puzzle is that philosophical linguists tend to divide words into form-words and content-words, into syncategoremata and categoremata. There is talk of a process called 'lexical insertion' as a separable part of the process of language-generation. But it is not clear whether 'I' and 'you' are lexically insertable items, to be classified along with 'Mary', 'John', 'chair', 'table', 'blue' and 'large' and to be distinguished from these as these are distinguished from one another, or whether they are more intimately related to language structure, as are 'of', 'is', 'and' and, most of the time, 'he', 'she' and 'it'. (The latter are sometimes demonstratives, usually mere place-holders like bound variables.)

Let us try to discuss 'you' and 'I' as a special case without raising the general issue of indexicals. Sentences divide, if we exclude other complications, into the *impersonal*, which do not alter their reference depending on whether you speak them or I, and the *personal*, which do. These changes of reference are immediately reflected in grass-root dialectics in our tendency to agree to impersonal sentences without paraphrase and to personal ones only after intersubstitution of 'I' and 'you'. Personal sentences include 'I am hungry', 'There's mud on your shoe', 'I Tarzan, you Jane'; impersonal 'Birds fly', 'It is raining', but also 'You and I are the same build', 'You were ten feet away from me', 'Your opinion is different from mine', in which 'I' and 'you' appear symmetrically. (A necessary and sufficient condition of symmetry appears to be the existence of a paraphrase using only 'we'.)

Let us suppose that in language-learning we can come to learn or decide that personal sentences fall into pairs  $\langle S_i, T_i \rangle$  such that you and I take ourselves to be in agreement about them when and, in general, only when I assert  $S_i$  and you  $T_i$ , or I  $T_i$  and you  $S_i$ . That is to say, we learn or decide that the pair of sentences  $\langle S_i, T_i \rangle$  has a pair of meanings  $\langle M_i,$

$N_i$  symmetrically related to us. The question then is: which, in which circumstance, means which? If, for example,  $S_i$  but not  $T_i$  contains 'I', how do I know that  $S_i$ , and not  $T_i$ , is what I must say to refer to me?

One suggestion is that, of course, I find out by *pointing*: 'I Tarzan, you Jane'. But for this to be possible it is first necessary that we understand the pointing gesture; and that, even if we understand it in application to impersonal contexts, we know whether and in what way it differs for personal ones. Custom might decree that personal pointing be effected by means of alternative and especially confusing gestures. And in any case, what if you are a Martian and we converse by radio-telephone?

Gesture and nuance (of course, you will say) have special intimacies, and one would *know* what was meant. But even so, and this is the point, how? By miming the equivalent of a personal sentence? Then the problem starts again. But one *must* in some sense mime the equivalent of a personal sentence. Hence to appeal to gesture without an account of how it is understood is a piece of wool-pulling. This cannot be the answer.

'I Tarzan, you Jane' was intended not, of course, to teach the personal pronouns, but to teach the names 'Tarzan' and 'Jane': we seem to need instead 'I I, you you'. But this, as we noted, is not even tautological: it is false.

One's next thought might be that, even if not private, pains, opinions and sense-impressions can be predicated of 'I' in ways and contexts in which they cannot be of 'you'. Although I may perceive that you are hungry or like Bach, it is not up to me (unless I am your analyst) to *tell* you of these things. And I can quote evidence of them, whereas you cannot. So that personal sentences do *not* fall into pairs, or not symmetrically. I know that 'I' is the first-personal pronoun of your language because you do not lightly deny my statements such as 'I think it is going to rain'. (Except in the sense that you may say you think otherwise.)

I have not bothered to state the more easily discredited argument from pure privacy, as distinct from an argument from relative privilege. Either might (might, not would) provide us with a successful solution if we could detach the question of the *meaning* of words such as 'think' and 'feel' from that of what subjects they are used with. Granted, for the sake of argument, that 'I think that  $S$ ' is privileged compared with 'you think that  $S$ ', how do I know that 'think' in both cases means 'think'? That 'I blorb' is more privileged than 'you blorb' does not imply that I am 'I', since 'blorb' might mean 'have a black spot on one's forehead'. Beside things I am better placed than others to discover about myself there are others I am *worse* placed to discover (and some my best friends do not tell me).

Then what about performatives, such as 'promise', that *transact business* between speaker and hearer? But speaker-hearer-business is always reciprocal and can be looked at from either point of view. 'I promise that X' means in context 'I promise *you* that X', and this is equivalent to 'You are promised X by me', which might just as well be abbreviated 'You are promised X'; or for that matter just 'X' is a promise'. Who promises, and to whom, is implicit in the act of speaking the promise, and performatives are not essentially personal at all. The same applies to point-of-order-raisers such as 'I don't understand', 'I didn't hear', and, perhaps, any personal locution whose point is to do with the conduct of the conversation to which it is a contribution. The argument might even be extended to 'think', 'feel'.

Do we learn 'I' and 'you' by overhearing them used by others? Certainly, I may understand how you use 'I' by hearing you use it not just to me but in the same terms to others; and by hearing others use it alike to me and to you. If I have a modest ability at induction this will do the trick for me. So the problem is solved.

But it would be very remarkable if this were the leading solution. It would have as a consequence that personal language would not be possible between just two people. Jane could not teach 'I' and 'you' to Tarzan, nor Robinson Crusoe to Friday; and for that matter a mother could not teach them alone to her family of infants, nor a direct-method teacher to his class of uniformly untutored foreigners.

Now it might very well be supposed that Jane and Tarzan *would* not use 'I' and 'you', preferring 'Jane' and 'Tarzan'; but that is a very different thing from supposing that they *could* not. For it is difficult to imagine any argument to this effect that did not tell equally against their ability to learn to refer to one another as 'Jane' and 'Tarzan'. Having established his pupils names, 'Itelo', 'Willatale' and so on, the teacher *can* perhaps teach 'you' and 'I' by simply announcing one morning 'You Itelo; you Willatale; I Henderson' and when someone echoes 'I Henderson', by frowning and saying 'You *Mtalba*, I Henderson'; and after this, he may establish Romilayu's name by hearing him say 'I Romilayu'. But these locutions are symmetrical ones, and teach a new mode of reference in terms of an old one already known. And why should we assume that impersonal reference is more natural, or more easily learnt, than personal? (If so—which it is possible to doubt—that would in any case be a psychological accident.)

So 'I' and 'you' must be learnable, after all, in rather the same way as names, by a rather simply inductive procedure or, as we might say, by coherence. I learn John's name, if I learn it in the primary way in which one learns names, by hearing, and eventually myself making, the noise 'John' in a variety of circumstances which for the most part cohere as contexts in which locutions about John are appropriate. In the same

way I learn 'I', if I learn it in the primary way in which I might learn it before learning names such as 'John', by hearing 'you'-locutions uttered, and myself coming to utter 'I'-locutions, in contexts for the most part appropriate to locutions concerning me. And the difference between 'I' and 'you' is the difference between 'Smith' and 'Jones', to which may be attached words predicating respective heights, build, hair-colours and also relative and absolute positions at certain times, or other words forming appropriate questions, greetings and so forth. I know that you mean me when you say 'you' because the predicates you attach to the 'you', context-dependent or not, indicative or interrogative or whatever, are usually predicates appropriate to me.

In this respect 'I' and 'you' are improperly named *pronouns*, and behave more nearly like names, and hence like lexical items or categoremata; or rather, they blur these distinctions. It would be possible to make some of the same points about 'now' and 'then', 'here' and 'there', 'this' and 'that'.

*University of New South Wales*

## THE PRESCRIPTIVIST DEFINITION OF 'BETTER'<sup>1</sup>

By M. B. SMYTH

'A is a *better* X than B' is to mean the same as 'If one is choosing an X, then, if one chooses B, one ought to choose A' (R. M. Hare, *The Language of Morals*, p. 184).

AS has been shown by M. K. Rennie ('On Hare's "Better"', *Nous*, vol. 2, 1968), Hare's definition is, in this form, open to fatal technical objections. The aim of the following paper is to find an improved version of the definition that will withstand Rennie's objections (as well as some other difficulties, to be mentioned below).

### I

Following Rennie, we symbolize the original definition—in a somewhat simplified form—as

$$(DB) \quad xBy = \text{df } Chy \supset OChx$$

Here '*xBy*' abbreviates '*x* is better than *y*', '*Chx*' abbreviates '*x* is chosen' (or 'the agent chooses *x*'), and '*OChx*' abbreviates 'the agent ought to choose *x*'. The 'field of choice' *X* and the implicit quantification over all agents are ignored, as being irrelevant in the present context.

<sup>1</sup> I am grateful to Professor Hare for some improvements in the exposition of this paper.

Among the simplest formal properties of the relation *better than* are that it is irreflexive and that it is transitive.<sup>1</sup> If the definition (DB) is to be regarded as adequate, the same must be true of the relation *B* introduced by it. However, it is easy to show that *B* has neither of these properties (here our presentation diverges slightly from that of Rennie):

(I) For *B* to be irreflexive would mean that, for any item  $x$ ,

$$\sim(Cbx \supset OChx).$$

A counterexample to this is obtained by letting  $x$  be any item which is not chosen; or any item which ought to be chosen.

(II) For *B* to be transitive would mean that for all items  $x, y, z$ ,

$$(Chy \supset OChx) \ \& \ (Chz \supset OChy) \supset (Chz \supset OChx)$$

A counterexample is provided by any triple  $(x, y, z)$  such that

$z$  is chosen,

$y$  is not chosen but ought to be chosen,

$x$  ought not to be chosen.

In particular,  $x$  and  $z$  may be the same item, giving the type of counterexample suggested by Rennie (p. 77).

A simple apparent remedy for the deficiency (I) (failure of irreflexivity) is available: assume that (DB) defines 'better-than-or-the-same-as' rather than 'better-than'; and construe ' $x$  is better than  $y$ ' as ' $xBy$  &  $x \neq y$ '. But this quickly proves inadequate. The relation *better-than-or-the-same-as* is not irreflexive, but it is surely antisymmetric. Now

(III) For *B* to be antisymmetric we must have, for all items,  $x, y$ ,

$$(Chy \supset OChx) \ \& \ (Chx \supset OChy) \supset x=y$$

Counterexamples are immediate: let  $x, y$  be any two distinct items, neither of which is chosen; or, let  $x$  be any item which is not chosen but ought to be chosen, and  $y$  any item distinct from  $x$ .

The conclusion is that Hare's definition fails to satisfy even the simplest of formal adequacy conditions.

How can we explain the fact that the definition is so easily vitiated by trivial counterexample? If the counterexamples are examined closely, it will, I think, be found that they all have a common source: in certain respects, it is too *easy* to make the definiens in (DB) true. We can make it true by taking as  $x$  any item which happens not to be chosen; likewise by taking as  $y$  any item which ought to be chosen—without requiring any significant relationship between the two items supposedly being compared. The definition entails that anything which is not chosen is worse than everything else; and that anything which ought to be chosen is better than everything else.

<sup>1</sup> For terminology of the logic of relations see, for example, P. Suppes: *Introduction to Logic* (van Nostrand, 1957).

The two components of the definiens,  $Chy$  and  $OChx$ , seem to be tied together too loosely. We might try to tighten the link by replacing the material implication operator by some form of strict implication; but, as Rennie shows, this approach is not very promising. A more adequate solution is suggested by recalling the intended (prescriptivist) meaning of 'O'. The force of ' $OChx$ ' is supposed to be to prescribe the choice of  $x$  and, moreover, the choice, by any agent, of any item relevantly similar to  $x$ . Now, the definition (DB) arbitrarily singles out the second item in the comparison for this "universal-prescriptive" treatment. This is quite unreasonable. It means, for example, that in the case of the two lecture-courses considered by Hare (ibid.), the tutor who tells his pupil that  $A$ 's lectures are better than  $B$ 's is saying something like: 'If you happen to choose to attend  $B$ 's lectures, then: You, and everyone else in a similar situation, are to attend lectures resembling  $A$ 's'. It is absurd that the choice by a particular student of a particular course of lectures should have such momentous consequences.

Clearly, the universal-prescriptive treatment ought to cover both subjects of the comparison. The verbal rendering should be something like: 'Whenever a choice-situation presents a pair resembling  $\langle A, B \rangle$ <sup>1</sup> let the first be chosen if the second is chosen'; (DB) should be replaced by

$$(DB') \quad xB'y = \text{df } O(Chy \supset Chx)$$

Let us examine (DB') from the point of view of the formal adequacy conditions for a definition of 'better-than-or-the-same-as'. The requirements of reflexivity and transitivity may be taken to be satisfied, provided we accept the deontic laws  $O(p \supset p)$  and  $O(p \supset q) \ \& \ O(q \supset r) \supset O(p \supset r)$ .<sup>2</sup> With regard to antisymmetry consider the statements

$$O(Chy \supset Chx), O(Chx \supset Chy)$$

Taken together they demand, in effect, that if either term in a comparison-pair relevantly similar to  $\langle x, y \rangle$  be chosen, then both terms must be chosen. This is unacceptable if  $x, y$  are distinct items one of which is better than the other; and it is also unacceptable if they are 'equally good', for in that case the agent should be permitted to choose either indifferently. The only alternative is that  $x=y$ : the requirement of antisymmetry is, indeed, satisfied. Thus, as far as concerns the formal requirements, (DB') seems adequate as a definition of 'better-than-or-the-same-as'. For 'better-than' we have

$$(DB'') \quad xB''y = \text{df } O(Chy \supset Chx) \ \& \ x \neq y.$$

<sup>1</sup> It is convenient to write simply ' $A$ ' instead of ' $A$ 's lectures', etc.

<sup>2</sup> Most systems of deontic logic run parallel to indicative logic at least far enough to have these as laws. A system based on Hare's prescriptivism would presumably have the same feature (cf. R. M. Hare, 'Some alleged Differences between Indicatives and Imperatives', *Mind* 1967).

## II

The example of the lecture-courses, mentioned above, is introduced by Hare in these terms: 'Now suppose . . . that I am asked by a pupil to advise him on the respective merits of several courses of lectures on Aristotle's *Ethics*'. But the example unfolds in terms of a choice between just *two* courses, *A*'s and *B*'s. This procedure involves an undue simplification of the problem. As will be shown in a moment, special difficulties arise in connection with choice-situations in which more than two items have to be compared.

Suppose, then, that the pupil has at least three courses of lectures from which to choose, and that these have been arranged by his mentor in (descending) order of merit *A*, *B*, *C*, . . . , *K*. Suppose that the pupil selects one of these courses—for definiteness, let us say *C*. Then it follows by a series of applications of (DB''), together with the prescriptivist analysis of 'ought', that he has rejected the advice in question unless he also selects all the preceding items in the list—that is, *A* and *B*. But this conclusion is unacceptable. This is most obviously true in case it is impossible to attend both *A* and *B* (say, because they are held at the same time of day). Even where there is no such hindrance, the pupil has not necessarily rejected the advice if he attends only *B* and *C*. For, although *A* is preferable to either *B* or *C* taken simply, it may be that the combination {*B*, *C*} is preferable to *A*, and even—because of the way in which *B* and *C* complement each other—preferable to either of the combinations {*A*, *B*}, {*A*, *C*}.

In searching for a more adequate formulation, it will be necessary to take account of certain features of the prescriptivist approach that we have so far neglected. In particular, we discuss the conditions of universalizability of judgments of comparative merit, or—what comes to the same thing—the descriptive content of such judgments.

What, then, is the "universal" implication of the judgment that, say, course *A* is better than course *B*? Is it simply that, if *A'*, *B'* are any other courses such that *A'* is relevantly similar to *A*, and *B'* is relevantly similar to *B*, then *A'* is better than *B'*? This is not, in fact, sufficient. Someone who asserts that *A* is better than *B* will normally be prepared, while maintaining the *same* criteria of comparative merit, to make the same judgment concerning many other pairs  $\langle A', B' \rangle$  not satisfying such a restrictive condition. If asked for the basis of his judgment that *A* is better than *B*, he will reply that the *A* lectures are more informative and more thought-provoking than the *B*, or something of the sort. Evidently this judgment generalizes to pairs  $\langle A', B' \rangle$  in which we cannot say that *A'* is similar to *A*, *B'* to *B* (e.g. the case already discussed, in which *B* is better than a third course, *C*; here  $A' = B$ ,  $B' = C$ ).

Thus the descriptive basis of a judgment of comparative merit lies, not in the qualities of either or both of the items being compared (taken



singly), but in an appropriate relation between them. More explicitly: ' $x$  is better than  $y$ ' implies (in some sense) that there is a descriptive relation  $R$  such that (i)  $xRy$ , and (ii) if  $x', y'$  are any other items for which  $x'Ry'$ , then also  $x'$  is better than  $y'$ .

The desired properties of the value-relation *better* are to be secured by making appropriate assumptions about the corresponding descriptive relations  $R$ . Two assumptions immediately suggest themselves: that  $R$  be transitive, and that it be irreflexive. But these will not be quite sufficient. A small complication arises from the fact that two items may be of equal value: we need to ensure (by means of assumptions about  $R$ ) that *equally good* be an equivalence relation. Also there is the question of comparability—i.e. the question whether, for any two items,  $x, y$ , it must be the case that either  $x$  is better than  $y$ , or  $y$  is better than  $x$ , or  $x$  and  $y$  are equally good. I do not want to enter into a detailed discussion of these points here, and will merely suggest one method of settling them: to stipulate that the relation defined by

$$\sim xRy \ \& \ \sim yRx$$

be transitive. Whatever may be the exact form of the assumptions made about the descriptive relations  $R$ , these relations will be referred to simply as 'orderings'.

Some attention must now be given to the prescriptive element in judgments of comparative merit. In this connection, we shall need to be more explicit than heretofore in describing the choices made by an agent. It will, for example, be necessary to take account of the relevant 'field of choice'. Again, on any given occasion on which the agent is choosing an  $X$  (i.e. choosing an item from the field  $X$ ), only some finite subset  $S$  of  $X$  will be actually available.  $S$  will be called the 'alternative-set', and we shall write ' $Cb(x, S)$ ' to indicate that the agent chooses the item  $x$  from the alternative-set  $S$ . We note also that, if  $S$  is an alternative-set included in a field on which the ordering  $R$  is defined, then  $S$  contains at least one  $R$ -maximal element, i.e. an element  $y$  such that for no  $x \in S$  do we have  $xRy$ . (This is trivial, since  $S$  is finite.)

The desired account of the 'prescriptive element' in judgments of comparative merit is as follows: A judgment of the form ' $x$  is a better  $X$  than  $y$ ', for which the relevant ordering is  $R$ , commits the speaker to prescribing that any agent, when making a choice from an alternative-set  $S$  included in  $X$ , should choose an  $R$ -maximal element of  $S$ .

This leads directly to a formal definition of 'better'. Since it is now necessary to make explicit reference to the field of choice, the formula ' $xBy$ ' is expanded to ' $xB_Xy$ '. The definition reads:

$$(DB_X): xB_Xy \text{ =df } \text{There is an ordering } R \text{ of } X \text{ such that} \\ (i) \ xRy$$

- (ii) for every alternative-set  $S \subset X$ , and all  $w, x \in X$ ,  $O(Cb(w, S) \ \& \ x \in S \supset \sim xRw)$ .

As with the original version, (DB), and its variants, 'better' is here defined in terms of 'ought'. On the present approach, however, this order of priority is inessential. Since clause (ii) is already purely universal in form, the definition can—with only a slight modification—be reformulated in terms of the universal imperative mood, instead of the operator 'O'.

The part played by the orderings  $R$  in  $(DB_x)$  may appear to give rise to a difficulty. For the definition is intended to be in broad conformity with Hare's analysis of evaluative terms. An important element in Hare's doctrine is that there is no (descriptive) property tied to the word 'good' as its meaning. Surely, by analogy, there ought not to be any descriptive ordering tied to 'better'?<sup>1</sup>

To resolve this difficulty, it is useful to recall Hare's distinction between meaning and criteria (*The Language of Morals*, Ch. 6). The orderings  $R$  embody the *criteria* by which, in particular contexts, comparative merit is assessed. It would be inappropriate to fix any one set of criteria as part of the meaning of 'better', for the reasons discussed by Hare in connection with 'good'. On the other hand—as is made particularly clear in section 8.2 of Hare's book—it is part of the meaning of an evaluative term *that there are* descriptive criteria for its application; and this is what  $(DB_x)$  is intended to capture. (A quantification over properties, similar to the quantification over relations in  $(DB_x)$ , occurs in Hare's definition of 'ought', p. 191.)

It is now easy to solve the problem discussed at the beginning of this section. The advice ' $A$  is better than  $B$ , and  $B$  is better than  $C$ ' has, according to  $(DB_x)$ , the following prescriptive implication: the pupil should, if he is choosing a single item from the alternative set  $\{A, B, C\}$ , choose  $A$ . Accordingly, he is to be regarded as having rejected the advice just in case he chooses  $B$  (alone) or  $C$  (alone). Nothing is prescribed as to the choice of a pair such as  $\{B, C\}$ , and the former difficulties do not arise.

Thus it appears that definition  $(DB_x)$  satisfies the technical criteria considered in this paper. It may seem that we have strayed rather far from the definition of Hare's with which we began. Yet  $(DB_x)$  is, I think, in better accord with Hare's *informal* treatment of commending and choosing (op. cit., esp. pp. 101–109) than  $(DB)$ – $(DB')$ ; it realizes the purposes for which Hare constructed his definition more adequately than they do.

<sup>1</sup> This point has been suggested to me by Professor H.-N. Castañeda.

## INVERSE DISCRIMINATION

By J. L. COWAN

THE justice or injustice of "inverse discrimination" is a question of pressing social importance. On the one hand it is argued that when a morally irrelevant characteristic such as race, creed or sex has been treated as morally relevant and injustices have resulted, it is then proper to treat that characteristic as morally relevant in order to make reparations. On the other hand it is argued that if the characteristic in question is morally irrelevant, its use even in this manner would still constitute discrimination, discrimination now in favour of those possessing the characteristic and against those not, but unjust discrimination still.

Public discussion of this issue all too rarely goes far beyond the level of the arguments as given. Yet the logic of these arguments is murky to say the least. It is therefore to be hoped that the analytical skills supposedly characteristic of philosophers might here play a valuable social role, and we are indebted to J. W. Nickel (*ANALYSIS*, Vol. 32, No. 4, pp. 113-114) for beginning such a clarification. I should like here to try to continue it.

Nickel maintains that the argument against inverse discrimination given above goes wide of the mark since the characteristic which is now operative is not actually the original morally irrelevant one. 'For if compensation in the form of extra opportunities is extended to a black man on the basis of past discrimination against blacks, the basis for this compensation is not that he is a black man, but that he was previously subject to unfair treatment because he was black . . . in such a case the characteristic which was the basis for the original discrimination (e.g. being a black person) will be different from the characteristic which is the basis for the distribution of special considerations (e.g. being a person who was discriminated against because he was black)' (p. 114).

The problem is that Nickel does not make it entirely clear just what he is about here. He may simply be pointing out that if a person has suffered injustice through morally unjustified discrimination, then reparation to that person will be appropriate. But surely it was not against this relatively uncontroversial point that the original argument was directed. And Nickel's formulation leaves open the possibility that he is actually trying to support the far more questionable claim which was the original target of that argument.

'Being discriminated against because he was black' is clearly a complex predicate. What I would like to suggest is that the portion of it which was morally irrelevant in independence remains so within the complex and is thus mere excess baggage. The reason why he was

discriminated against is not what should now ground reparation, but rather simply the fact that, and extent to which, he was unjustly discriminated against for whatever reason. Thus, assuming that the discrimination is otherwise the same, we would presumably not wish to say that Jones, who has been discriminated against as a black, should now be favoured over Smith, who has been equally discriminated against as a woman or a Jew or whatever. We are therefore left without a moral relevance for blackness, and thus without a moral basis for inverse discrimination based on blackness as opposed to discriminatory injustice *per se*.

Nickel's reasoning thus does not really, as it might be taken to do, provide any support at all for the kind of self-contradictory thinking the original argument was surely intended to rebut. This is the reasoning that since blacks, to retain this example, have suffered unjust discrimination we should now give them special treatment to make it up to them. Once again there is no problem insofar as this simply means that where individual blacks have suffered injustice it should, as with anyone else, insofar as possible be made up to those individuals who have so suffered. The fallacy arises when rather than individuals it is the group which is intended, and individuals are regarded merely as members of that group rather than in their individuality. This creates a contradiction since the original premise of the moral irrelevance of blackness on the basis of which the original attribution of unjust discrimination rests implies that there is and can be no morally relevant group which could have suffered or to which retribution could now be made. Thus those who would argue that since "we" brutally kidnapped "the" blacks out of Africa and subjected "them" to the abominations of slavery, or that since "we" have exploited and degraded "women" since Eve, "we" therefore now owe retribution to our neighbour who happens to be black or a woman, are involved in inextricable self-contradiction. Except to the extent he or she as an individual has unjustly suffered or will unjustly suffer from this history while we as individuals have unjustly profited or will unjustly profit there can be no such obligation.

Nickel's original formulation is thus ambiguous. 'The context in which this difficulty is likely to arise is one in which a group has been discriminated against on the basis of morally irrelevant properties, but in which this discrimination has been recognized and at least partly come to an end; and the question at hand concerns how the members of this group should now be treated. Should they now be treated like everyone else, ignoring their history, or should they be given special advantages because of past discrimination and injustices?' (p. 113). Once the question is disambiguated the answer is clear. They should most certainly be treated like everyone else. But this does not mean "their" individual histories should be ignored. As with anyone else, injustices

moral point of view will lead to a miserable life for the person who so acts. In short, there are solid self-interested reasons for a man not to override moral considerations. Indeed, *only* to be a man of good morals is not, as Kant stressed, to be a morally good man. But there are generally good self-interested reasons for becoming genuinely unselfish and for becoming a moral agent (p. 318).

There is no contradiction here. What Nielsen means is this. Where a person has to choose between acting rightly or acting self-interestedly in a specific case, to urge him to do what it is right to do will have no effect unless moral considerations matter to him. He will not appreciate your appeal if they do not; but then how is one ever to establish independently of moral concern that to do what is right here is what it is best to do? But where the case is not one of a choice which confronts the individual now, a different sort of consideration (we are told) arises: the consideration of what a man should do, not now, but *throughout his lifetime*. The important paragraph (which I have kept to the last) is this one:

While there are, or at the very least may be, specific situations in which we can find no objective reason for acting as a moral agent rather than as a man who is looking out for number one and only number one, we must take care not to draw the mistaken conclusion that this shows that there can be no reasons of an objective sort for adopting, as an overall life policy, a moral point of view, as distinct from a purely egoistic point of view. It is one thing to say that there are some situations in which doing what is right will not bring one happiness and will not serve one's rational self-interest, but it is another thing again to claim that a man who made egoism the maxim of *all* his actions would be happy or would, by consistently acting on this maxim, in reality serve his own self-interest. That this would be so is very questionable (p. 317).

The idea is: reasons can be given a man which will incline him *to adopt* the moral point of view; he is to be given 'reasons (cf. above) for becoming genuinely unselfish'—'for becoming a moral agent'.

There are a number of things in these passages about which I have difficulty, but I should like here to attend only to one of them.

There is a distinction drawn within any moral tradition between the man who is a good man and the man who seems to be a good man. The good man is a man who cares how his life goes, *in a certain way*. He is concerned to do what it is right to do. But the evil man is sometimes concerned to do what it is right to do as well; when it is to his advantage to do what it is right to do, as when, for example, he wishes to present himself (out of fear, or cunning) as someone who does right. But his concern here is not like the good man's concern. He does not care *to be* good. He cares to appear to be good. And while he does here what the good man does, at the same time he does not do what he does. The good man acts out of a regard for what it is good or right to do. The evil man acts with a regard to what right requires, but *out of* a regard for what he

wants. Unless one admits to this I do not see how the distinction between the good man and the dissembler can be made out. Nor do I see how one can account for the importance in morals, and therefore in our lives, of the character of a man. Morally regarded, what you do when you act is not just a matter of what appears to be being done. For all its appearance, if it is not done out of a genuine goodness of intention and concern, then for all that it is not a wrong act, it is not a good act either.

'To do what is right' can mean either: to do what the good man does; and that is to act out of a concern for right. Or it can mean: to do what a good man would do, whether or not one cares to be good.

It will not do to say that if a man does good actions all his life then for all intents and purposes he is good. For what does the expression 'does good actions' mean here? If it has the first of the two senses above there is no quarrel (and no need for 'for all intents and purposes'). If it means 'He outwardly performs good actions, he shows no signs of evil' then the question 'But is he a good man?' has a place. And do not say that if he *never* shows an evil will or badness the question is pointless. That we can (*ex hypothesi*) get no answer does not mean that the possibility does not obtain; being good is not just a matter of how you "act"—for no matter how you act the possibility that you are not a good man is not exhausted. It is not as if a lifetime of good actions, publicly, or even privately, performed, created a vacuum in which this possibility expired. This is not so, for the possibility is not a possibility about your actions (in that sense) but about *you*. And you are there all along.

Now if 'do what is right' is ambiguous, so too is 'take the moral point of view'. In a sense Dickens's Uriah Heep takes the moral point of view. He is careful always to do what the English moral tradition requires—when he is on exhibit. On any issue we could ask Uriah, 'Well, what point of view will you regard this matter from in order to decide on a course of action?' And sometimes he will say: 'From the moral point of view you dimwit. Don't you think I want to get on?' To act humbly is something you can do on a stage, or in the face of your betters. *To be* humble—Is that something you can do on a stage? Is that something you can *choose* to do?

Nielsen speaks, in a passage already quoted, of *adopting a moral point of view*, as distinct from a purely egoistic point of view. These words are his own, and 'purely egoistic' can only be a pleonasm. Earlier in that work, after remarks about the kind of 'system of social control' morality is, Nielsen remarks:

I am talking of the man who is resolved to be guided by reason, who has what has been called a 'commitment to rationality'. Can we show this individual that simply considering his own point of view, if he is to act rationally, the best thing for him to do is to adopt that moral point of view. . . .?

I have already tried to establish that, considering society at large, we should adopt such a system of social control [as morality is]. But the question I am now asking is a different one. It is a question about what is the most rational thing for an *individual* to do when he is trying to decide what order of priorities are to count most heavily *for him* in the conduct of *his life* (pp. 314-315). (Nielsen's emphasis.)

What concerns me here is the notion that a person is *to decide* what order of priorities are to count most for him in his life. We are to believe that one among the possible decisions here is: moral considerations are to count most with me. It is this decision that is to count as adopting the moral point of view.

Now clearly you can adopt the moral point of view in the sense in which Uriah Heep adopted it. You can on any occasion choose to do what a good man would do. If you don't know what a good man would do you just inquire and then do it. And I suppose a man could at some point in his life decide to do this and keep to it until his death. But would his have been a good life, morally regarded? Would he have adopted *morality*?

What would make his life a good life, morally considered? Would it not be, not merely that he never did a wrong act, but that he loved the right? By 'loved the right' I mean, it mattered to him that he did right. It mattered, not because of something it got him, or because he had decided to do it, but because it was right to do it. But that is to say he had a regard for goodness, and can you *decide* to have a regard for goodness? In time he may come to have this regard, but until he does his life is different from that of the good man, and how is he to decide that it shall be the same? A regard for goodness is a certain kind of concern with life, what people call a moral concern. How is this concern to arise? A man can decide to concern himself with something. He decides, say, to consider and deal with it in a certain way. You can concern yourself with, for example, other people and their problems in this way. You listen, hear, and take action. You get involved. But it is possible to be involved only so far. And it is the kind of involvement that determines whether there is moral concern here. We distinguish caring about and using people. And we distinguish between caring and just being involved. But can you decide to care?

If honesty matters to you, or selflessness, it matters. But if it doesn't matter to you, can you make it matter? I am not saying, if a man does not care about being honest he will never come to care. He may come to care. But can he at any time decide to care? Can you choose to be an honest man in the way the good man is honest?

I am afraid I cannot get a sense for the expressions 'take the moral point of view', 'adopt morality', except that in which Uriah Heep took

or adopted it. He adopted it in the sense in which a man can adopt a policy one day, and periodically discard and re-adopt it thereafter.

The trouble is, you cannot *discard* morality.<sup>1</sup> Shame is proof of that, and remorse, and the guilt which brings men to confess who would never be discovered. If a man who has a sense of right and wrong commits a wrong, he may come in time to wish he had not done it. He may wish this, not because of some consequence of that action which now plagues him, but because of the action it was. It is because of the wrong it was that he regrets, and suffers, as he dwells on what he did. If you say the suffering is a consequence of the action, it is, but this is not to the purpose; for this consequence, his suffering, is internally related to what the action was. It is his appreciation of the wrong he did that makes him suffer. And what he wishes is not that he should cease to suffer, but that he should never have done it.

If he does wish that he should cease suffering, this too is instructive. For it points up what I meant when I said you cannot discard morality. If he suffers, because of his sense of the wrong done, the evil he was capable of, why does he not simply discard morality and suffer no more?

How is a man to adopt the 'moral point of view'? How is he to *decide* to be kind, unselfish, courageous, honest, loyal, just, and so on? For that is what 'having the moral point of view' is. The moral point of view is not some point *above* honesty, selflessness, mercy, integrity, justice, in terms of which those are seen to be worthwhile. Having the moral point of view is just approaching life and persons honestly, justly, forgivingly, and so on, where these things matter to you, where, in a sense, these things are what you are. But if you don't have a regard for, say, honesty now, how are you to decide to have one? My trouble is not: I cannot see what sort of thing would be a reason for deciding. My trouble is: is this the sort of thing you can decide?

'We can always ask for a justification for taking a moral point of view at all.' 'Why should I be moral?'

The important words are 'be moral', 'take the moral point of view'. The 'reasons for being moral' are of course to be reasons for being moral, not just seeming to be. The enterprise is to present you with reasons which are to determine your choice, where this choice is to be moral, to take the moral point of view. But even if reasons are given which you own to be good, what then? You are to *decide* to be moral?

It is, I submit, the apposition of 'decide' and 'be moral' that generates the absurdity here. It is, if I might put it this way, not so much that the question 'Why should I be moral?' is absurd, as that what is supposed to *follow* it is absurd. You cannot decide to be moral.

<sup>1</sup> In this sense you cannot *adopt* the *self-interested* point of view.



## ON GIVING REASONS FOR BEING MORAL

By KAI NIELSEN

RODGER Beehler<sup>1</sup> is surely justified in stressing the importance of the Kantian distinction between a man of good morals and a morally good man. Both men, though with different attitudes in mind, will be concerned to do what is right. But the morally good man will do it out of a regard for what is right, because he believes it to be right, while the evil or amoral man will do what is right only because he sees or thinks he sees some advantage in it. But it is not the case that a following out of the implications of this distinction undermines either my claim that an individual can sensibly ask why be moral (why do what he acknowledges is morally required of him) or my companion claim that the moral point of view is something that could be adopted or rejected, i.e., that a man might decide not to take moral considerations as overriding all other considerations in the context of action.

The claim is made by Mr. Beehler that it makes no sense to ask such questions because he believes that it makes no sense to speak of *deciding* to allow moral considerations to count most with one, or to speak of deciding to be just or honest or kind. In general one cannot decide to be moral or to discard one's commitment to taking moral considerations as overriding all other considerations. There is, according to Beehler, no deciding to have a regard for goodness; there is no way of deciding to care for others. There is no deciding to be moral or immoral.

On the contrary, I think there is a tolerably straightforward sense in which an individual can decide to be moral. What this comes to here is a resolving to be moral. It is not that honesty or selflessness just matters to you or it does not. They can matter to an individual; and yet there are times when he is not honest or selfless, and as an intelligent agent with interests, wants and needs, he looks about him and he sees that people often are—and frequently to their advantage—dishonest and anything but selfless, and he sees honest and selfless men harmed.<sup>2</sup> He sees all that; he sees the defeat and repression of progressive social movements and the constant spectacle of evil and of human indifference. Faced with this, he may in a certain mood—indeed perhaps in a despairing mood—ask if there is any rhyme or reason to this concern of his with selflessness and honesty. He may very well come to feel that to live in this way is an arbitrary “act of faith” on his part, and that it is the case that he and others like him who are so concerned are being duped and used. He may out of his disillusion and resulting cynicism decide that moral concern

<sup>1</sup> Vid. sup., pp. 12–16.

<sup>2</sup> No sophisticated talk about the way(s) in which a good man cannot be harmed will suffice to undercut the force of the plain fact that there are still evident enough ways in which he can be harmed.

should not really matter so much, should not play such a central role in his life, and he may resolve to put a concern with them aside, except in the purely instrumental way that a Uriah Heep may be concerned with them.

For some human beings, by contrast, it might go the other way. They might be self-interested creatures reasonably hardened to others. But human beings are often not of a piece and they might come ambivalently to feel that there might be something in having moral concern and developing selflessness, and they might strive to make these things matter to them or at least wonder whether, after all, they really should make these things matter to them. And they might succeed in making these things matter. There is no conceptual ban (*pave* Beehler) on their trying or resolving to make them matter and indeed in their even making them matter.

In the first instance we have a man quite intelligibly wondering whether he should discard the moral point of view, and in the latter instance we have a man wondering whether he should act in accordance with it, as a morally good man—a man of good will—would act in accordance with it.

It is at best a psychological truth and not a conceptual truth that shame, remorse and guilt will accrue because of a negative decision in the latter case or a positive decision in the former case. But even if there were such psychological reactions, they would not show that a man cannot reject the moral point of view, for he might no longer attach *authoritative* weight to those feelings but might come to regard them as irrational hang-ups to be overcome. (Think of a man who might come to think of his conscience simply as his superego.) Most of the main characters in *No Exit to Brooklyn* had once, at least in some rudimentary way, learned moral notions, but they dispensed with them, discarded them, in the way many people discard religious beliefs.

Men who are not kind or just or loyal or honest can decide to *try to become so*. That is to say, they can so resolve. This can be as a result of deliberation, and in so deliberating they can decide to adopt a moral point of view—a point of view in which they will try to take a proper regard for considerations of honesty, selflessness, mercy, integrity and justice. And since immoral or amoral men, or just torn, ambivalent men, can by their deliberate decisions come to adopt the moral point of view, moral men can by deliberate decisions opt out of trying to live in accordance with the dictates of morality. Smerdyakovs are not conceptual impossibilities.

Which point of view a man adopts is a *logically* contingent matter and thus it is a real question and not a pseudo-question whether an individual should be moral. We know what it would be like for him to go one way and we know what it would be like for him to go the other, and we want

to know how, if at all, he could justify (without begging the question) going on in one way rather than another. It is not, however, evident what such a justification would look like. But it is the case that a moral point of view and various non-moral points of view are—as far as conceptual possibilities are concerned—open options for him.<sup>1</sup> He wants to know if any objective, decisive or even nearly decisive, but non-question-begging, reason can be given for his adopting one point of view rather than the other.

That we do not know what would count as an answer here does not show that such a question is a pseudo-question. We would only know that it was a pseudo-question if we knew that nothing could, logically could, count as an answer to such a question. Beehler has not shown that this is so and I know of no sound argument for believing it to be so. To meet Beehler's criticisms, I have simply set a context (one of several contexts) in which 'Why be moral?' could be asked without invoking incoherent or thoroughly problematic notions. Thus the perplexing question 'Why be moral?' remains with us. And to acknowledge that it is a genuine question, though indeed a very perplexing one, is not to give to understand that one thinks it has no answer.<sup>2</sup>

<sup>1</sup> It isn't just that morality can conflict with self-interest, it can conflict with various forms of privilege and with family, class and tribal considerations as well.

<sup>2</sup> In my 'Why Should I be Moral?', *Methodos*, Vol. XV (1963), I try to give the kind of answer that I think can be given to such a question. This essay has been reprinted in Kenneth Pahl and Marvin Schiller (edd.), *Readings in Contemporary Ethical Theory* (Englewood Cliffs, New Jersey: Prentice-Hall Inc., 1970).

*University of Calgary*

## MORALS AND REASONS

By RODGER BEEHLER

KAI Nielsen gives what he claims are two instances of deciding to be (or not to be) moral.<sup>1</sup> It turns out that these are instances of *resolving* to be, or not to be, moral.

In the first case a man resolves not to let moral considerations count with him. He 'decides that moral concern should not really matter so much, should not play such a central role in his life'.<sup>2</sup> He resolves to put moral concern aside. Just before that we are told that 'he may very well *come to feel* that to live this way is . . . arbitrary', and so on. Now this last may indeed happen to a man. The question is whether he can decide that it should. Earlier we are told that it is in a moment of despair that

<sup>1</sup> Vid. sup., pp. 17–19.

<sup>2</sup> Is 'so much' a hedge? Are morals to have *some* 'role'?

his question about the worth of his being honest or selfless arises. Now a feature of despair is that one despairs *of something*. And how a man comes to despair is often instructive. It shows us what really mattered to him. Kai Nelsen suggests that why this man despairs is because a lot of people who care nothing for living rightly get on very well. But why should he *despair* at this? A good man, reflecting on this fact, *might* despair. He might despair of human lives, and our treatment of each other, ever being as they ought to be. But why should he despair of the *importance of morality*, of honesty, forgiveness, mercy, loyalty, and the rest? If the fact that selfish and uncaring people get on disturbs his sense of the importance of living rightly, then maybe that is because he sees living rightly as something that ought to issue in something else. But then he doesn't despair of morality, but of the policy of "being moral".

In any case, if he *comes* to feel that living rightly doesn't matter, well then (perhaps) it doesn't matter now to him. The proof of that will be in what comes after. But *before* he feels that can he decide to feel that way? Can he decide that it won't matter? I do not deny that he might *resolve* not to attend to or appreciate the force of moral considerations in the future. He decides to be hard as nails. But to decide not to care is not like, say, deciding not to take care. You can (ordinarily) take care, or not take care, as you like. But your resolving not to care does not end your caring as your decision not to take care ends your taking care. Perhaps you will still care even when you wish you did not, when you wish that the result of that caring was something you were rid of.

If deciding not to be moral is resolving *to try* not to be moral, then a man could, I suppose, decide to do this. He could decide to try not to appreciate things morally. Certainly he can decide never to do what is right. But not to do what is right is not the same as not to appreciate things morally. And while you can decide to do the first you can at best decide to try to do the second. And your attempt may fail.

Nielsen's second case is one of 'deciding to be moral'. In the opening paragraph of his reply he repeats his claim that 'an individual can sensibly ask why be moral (why do what he acknowledges is morally required of him)'. Now again: 'Why do what he acknowledges is morally required of him?' is ambiguous between (1) 'Why do what he accounts he should do?' and (2) 'Why do what others who are moral account he should do?' If he admits that he *should* do the thing then why does he ask why he should? How does he get a question going here?

Nielsen's example is of a person or persons '*reasonably* hardened' to others, but who '*come ambivalently to feel*' that there might be something to moral concern, and who '*might strive to make* these things matter to them'. The important expressions I have italicized. I have nowhere denied that a man might come to be someone for whom living rightly

135 A.H. 19

matters as it does to the good man. He might become that. But he might not. And at no time can he decide to become like that—unless this is like deciding to become appreciative of great music. He can decide to try to appreciate it. And *perhaps* he will succeed.

A man can strive to become better. Ordinarily the striving gets its sense (certainly for him) from an appreciation he already has of the importance of living rightly, of being honest, unselfish, caring. A man without moral concern can, Nielsen tells us, strive to become a man with moral concern. I have only two remarks here. (1) How does he know *at what* he strives? How does he know how the good man's life is if he doesn't now appreciate moral considerations himself? (2) In Nielsen's example this problem is covered over by the expression 'come ambivalently to feel'. But (a) that he is ambivalent *means* that he is now drawn both ways. To be *drawn* is already to see, to appreciate, something in what you are drawn to. (Unless what you are drawn to is, say, the admiration it wins. Then we are returned to self-interested regard; though even here, as Peter Winch<sup>1</sup> has shown, there will often be appreciation, but corrupted, of what one wishes to be admired for.) (b) That he can *come to feel* (to appreciate) that there is something in morals is not a possibility I have denied. All I have asked is: can he decide to appreciate it that way?

<sup>1</sup> *Moral Integrity* (Basil Blackwell: 1968), pp. 24–5.

University of Victoria

## 'BEDEUTUNG' IN FREGE: A REPLY

By V. H. DUDMAN

IN 'The Meaning of "*Bedeutung*" in Frege' [ANALYSIS 30.6 (June, 1970), pp. 177–189] Professor Ernst Tugendhat takes exception to the Geach-and-Black rendering of '*Bedeutung*' as 'reference', claiming that it misleadingly suggests that 'what Frege meant by the *Bedeutung* of an expression is the object which the expression names' [p. 177]. It turns out, however, that Tugendhat's quarrel is not so much with Frege's translators as with Frege himself; he is bent less upon amending our English version of '*Über Sinn und Bedeutung*' than upon revising the original article.<sup>1</sup> Among other things he seeks to revoke Frege's doctrine that sentences are names of objects. Tugendhat's proposed modifications of Frege I find initially obscure, particularly as regards their ontological

<sup>1</sup> First published 1892. Translated as 'On Sense and Reference' in Geach, P. and Black, M., *Translations from the Philosophical Writings of Gottlob Frege*, 2nd ed., Oxford, 1960; pp. 56–78.

P, 6339

import. Once elucidated, they are found to represent a far more radical revision of Frege's ontology than Tugendhat has led us to expect. Moreover, they directly entail the renunciation of Frege's theory of predication: in view of the title of Tugendhat's paper, this result comes as a surprise.

Now some one might deny that 'the *Bedeutung* of an expression is the *object* which the expression *names*' either because he thought (i) the *Bedeutung* of an expression is (sometimes) something other than an *object*, or because he thought (ii) an expression does not always *name* its *Bedeutung*, or even, perhaps, because he thought (iii) there need really be *no such thing* as the *Bedeutung* of an expression. It is a defect of Tugendhat's discussion that he nowhere distinguishes propositions (i) and (ii), and indeed seems to conflate them. Thus the central paragraph of p. 182 shifts back and forth between speculating why Frege 'applied the terminology of the name-relation to the [*Bedeutung*]<sup>1</sup> of sentences' and trying to 'account for Frege's conception of the [*Bedeutung*] of a sentence as an object'. See also the central paragraph of p. 179, and the opening paragraph of the article. The distinction between propositions (i) and (ii) needs observing because it is at any rate arguable that (i) is incoherent in a way in which (ii) is not. Thus Frege's own doctrine (which is an instance of (i)) that predicates have *Bedeutungen* which are not objects gives rise to the antinomy familiar as 'the concept *horse*'. I locate the source of this antinomy in the tenet that there are entities which are not *objects*, in the broad Frege-Quine sense of 'object'. According to Frege's suggestion, an object is that which can be designated by a proper name. What (i) asserts is that some expressions have *Bedeutungen* unfitted to be values of variables occupying singular-term position, and in my 'The Concept *Horse*' (*Australasian Journal of Philosophy* 50 (1972), pp. 67-75) I have maintained that this proposition is not finally intelligible. If the *Bedeutung* of an expression is anything at all, it is an object. By contrast, the doctrine (ii) that an expression need not *designate* its *Bedeutung* is perfectly intelligible. There is nothing incoherent about the claim that an expression for a certain entity is yet not a proper name of that entity: naming is not the only semantic relation. Thus it is customary at once to allow that predicates determine classes as their extensions and to deny that predicates are names of classes. Again, certain theories postulate a systematic correlation between closed sentences and "propositions": a sentence is said to *express* its proposition—but to express is not to name. Equally, then, it is tenable that the *Bedeutung* of an expression, though admittedly an object, need not be an object *designated* by that expression.

Still, in what follows that distinction will be less important than the ontological gulf between (i) and (ii) on the one hand and (iii) on the other. Let me therefore underline this antithesis. According to the

<sup>1</sup> Tugendhat translates this as 'Significance'. I retain the original.

first two, the *Bedeutung* of an expression is an *entity* of some kind: to say that an expression has a *Bedeutung* is to say that *there is* something which is the *Bedeutung* of that expression. According to the third, by contrast, discourse about *Bedeutungen* can be ontologically innocent, and in particular the affirmation that a certain expression *has* a *Bedeutung* may be merely a *façon de parler*, claiming existence for no extra-linguistic correlate of that expression. What is envisaged in (iii) is some sort of contextual paraphrase: discourse ostensibly about the *Bedeutungen* of (at least certain sorts of) expressions is channelled down through a fixed number of contexts, and each of these contexts is translated as a whole into terms which require the existence of no such entities.

Which of (i)–(iii) provides Tugendhat with his ground for denying that the *Bedeutung* of an expression is the object which the expression names? To determine this, let us examine his proposed interpretation of 'Über Sinn und Bedeutung'. According to 'the correct account' [p. 183], 'what Frege himself really meant' by the *Bedeutung* of an expression is its *truth-value potential* (TVP), explained as follows [p. 180]:

- (1) Two expressions  $\phi$  and  $\psi$  have the same TVP iff, whenever each is completed by the same expression to form a sentence, the two sentences have the same truth value.

This explanation is less clear than Tugendhat seems to think, for we can construe (1) in two quite disparate ways. On the one hand we can read it straightforwardly as laying down criteria of identity for certain *entities*, viz. for the TVPs of expressions (without of course saying what manner of entities these TVPs are). As follows:

- (1a) The TVP of  $\phi$  = the TVP of  $\psi$  iff, whenever  $\phi$  and  $\psi$  are completed by the same expression to form sentences, these sentences have the same truth value.

Alternatively, (1) can be regarded as a *definition*,<sup>1</sup> as defining a two-place predicate of expressions, viz., the predicate 'are alike in TVP':

- (1b) ' $\phi$  is like  $\psi$  in TVP' for 'whenever  $\phi$  and  $\psi$  are completed by the same expression to form sentences, the sentences have the same truth value'.

The sole conceivable motive for introducing such a predicate by definition, it seems, would be in order to deny the existence of TVPs. (Hence interpretation (1b) is incompatible with interpretation (1a), where the talk is of actual objects, viz. TVPs.)

Which construction does the author himself place on (1)? The

<sup>1</sup> On p. 180 Tugendhat describes it as 'a general definition for the truth-value potential of an expression', but of course (1) cannot be construed as an explicit definition of 'the TVP of  $\phi$ '. A little later (pp. 180–181) he speaks of it as a definition of *sameness* of TVP, thus apparently endorsing my interpretation (1b).

question is urgent because Tugendhat's leading thesis is that the *Bedeutung* of an expression is its TVP. Under interpretation (1a), therefore, *Bedeutungen* are objects (they are entities concerning which the question of differentiation arises). But on interpretation (1b) there is really no such thing as the *Bedeutung* of an expression. The answer to our question seems, alas, to be that Tugendhat understands (1) now in one way and now in the other (without, however, acknowledging that he is tacking to and fro between *incompatible* interpretations) depending upon whether  $\phi$  and  $\psi$  are *predicates*, or *complete sentences*, or *proper names*.

First, predicates. Here Tugendhat opts unequivocally for (1b), and even portrays Frege himself as sharing the view that the *Bedeutungen* of predicates are really fictions. He cites a passage from the *Nachlass* in which Frege (most guardedly, for fear of "the concept *horse*") allows that two predicates can be said to agree in their *Bedeutung* iff they are coextensive, and remarks [p. 185]:

So here we have a case where Frege refuses to consider the [*Bedeutung*] as any kind of object whatsoever and is content to say under what condition two predicates have the same [*Bedeutung*]. And although Frege does not put it in these words, this condition clearly consists in having the same truth-value potential.

A little later, in the same vein, he disparages 'the usual account, according to which . . . concepts are regarded by Frege as "entities"' [p. 185]. When applied to "concept-words", then, (1) is treated by Tugendhat as simply providing a contextual paraphrase for the indiscernible predicate 'has the same *Bedeutung* as'.

So also for complete sentences: here again Tugendhat interprets (1) as (1b). For him, as for Frege, two sentences have the same *Bedeutung* iff they have the same truth value [p. 180]; but whereas Frege conceives of sentences as proper names and takes their [*Bedeutung*] to consist in one of two objects, "the True" and "the False" [p. 178], Tugendhat himself rejects 'Frege's application of the terminology of "name" and "object" to sentences' [p. 182]. He disbelieves in the *Bedeutungen* of sentences: to say that sentences  $\phi$  and  $\psi$  have the same *Bedeutung* is for him to say no more than that  $\phi$  and  $\psi$  are either both true or both false. In short, it is reason (iii) which prompts Tugendhat to deny that the *Bedeutung* of an expression is the object which the expression names—at any rate when the expression in question is either a complete sentence or a predicate.

How does he interpret (1) for singular terms? He assures us [p. 181] that if we take the *Bedeutung* of an expression to be its TVP, 'it then turns out, instead of being assumed, that the [*Bedeutung*] is, in the case of names, the object referred to'. In this case, then, he embraces interpretation (1a): the TVPs of proper names are the actual objects named. But why does it so 'turn out'? Obviously the result does not follow from (1a) alone, for (1a) does not tell us what TVPs *are*, but only when they are



identical. If we are to conclude on the basis of (1a) that the TVP of a singular term is its designatum, we need to be assured independently that singular terms interchangeable without loss of truth in closed sentences coincide in *nothing else besides* their designata. And Tugendhat is able to offer us no argument for supposing *that*. He does indeed remark that when  $\phi$  and  $\psi$  are proper names, (1) becomes 'only a cumbersome way of expressing . . . Leibniz' Principle of the Identity of Indiscernibles' [p. 180], but this is a mistake. The nearest thing to (1a) that the Identity of Indiscernibles can be taken as affirming is that two proper names have the same *designatum* (*not*: the same TVP) iff they are intersubstitutable *salva veritate*. Tugendhat's conflation of (1) with the Identity of Indiscernibles is thus an outright *petitio*, amounting as it does to an identification of the TVP (i.e. the *Bedeutung*) of a name with its reference. Apparently Mr. Dummett has already called attention to this fallacy: a footnote acknowledges that 'it is not strictly correct to say that the truth-value potential *is* the object referred to. All we can claim is that two names that refer to the same object have the same truth-value potential'. But Tugendhat seems not to realise the straits he is in, for he continues calmly: 'Consequently it would be preferable to say that the truth-value potential of a name is, rather than the object referred to, its reference to that object' [p. 181n]. Where does this leave us? Tugendhat's thesis, as has been said, is that the *Bedeutung* of an expression *is* its TVP. He now concedes that the TVP of a proper name is not the object it names.<sup>1</sup> It follows that he has disestablished Frege's rock-bottom usage, according to which the *Bedeutung* of a singular term *is* the object it designates. For this result we have been left totally unprepared: after all, Tugendhat is tendering his TVP as 'what Frege really meant' by '*Bedeutung*'. But if anything is certain it is that, for Frege, 'the moon itself is the *Bedeutung* of the expression "the moon"'.<sup>2</sup> Even as an amendment of Frege's doctrine, Tugendhat's proposal is unacceptable: no semantics can claim to be remotely Fregean which does not start from the proposition that a proper name's (*gewöhnliche*) *Bedeutung* is its designatum.  
reference

Nor does Tugendhat's violence to Frege end there. We have seen that he accepts proposition (iii): in particular he denies that there is any such thing as the *Bedeutung* either of a predicate or of a sentence. In the case of predicates he even ascribes this tenet to Frege himself. But the suggestion that Frege meant his talk of concepts as an ontologically

<sup>1</sup> The TVP of a name is not the object it names but 'its reference to that object'. It does not matter how we understand this enigmatic admission. Whether we continue to read (1) as (1a) and think of a name's reference to its designatum as a new object, distinct from the designatum itself, or whether more plausibly we construe the admission as a retreat to interpretation (1b), the upshot is the same: either way, Tugendhat's account of TVP diverges fundamentally from Frege's account of *Bedeutung*.

<sup>2</sup> Frege, G., *Nachgelassene Schriften*, ed. Hermes et al., Hamburg, 1969, p. 275.

innocent figure of speech will not withstand exposure to his writings:<sup>1</sup> Frege's repeated struggles with the antinomy known as 'the concept horse' are inexplicable except on the assumption that he was trying to assert that *there are* "unsaturated" *entities* which predicates have as their *Bedeutungen*. The case of statements is different in as much as Tugendhat of course acknowledges that Frege did not share his disbelief in the True and the False. The notion of TVP, intended to accommodate Tugendhat's own persuasion that 'the [*Bedeutung*] of a sentence cannot be thought of as an object named' [p. 179], is in this case consciously proffered as 'an *alternative* account of [*Bedeutung*] to that given by Frege' [p. 181; my emphasis]. Rejecting the view that sentences are names as an insufficiently motivated doctrine [p. 182], Tugendhat altogether fails to appreciate that he thereby abrogates the most distinctive feature of Frege's semantics, his *functional theory of predication*. Although the doctrine that sentences are names of objects was not announced by Frege until 'Function and Concept' (1891), it was implicit in his work from the beginning. For it is entailed by the treatment of sentences as completed functional expressions, and that treatment dates from *Begriffsschrift* (1879).

In Boole, predicate logic (the logic of "categoricals") and statement logic (the logic of "hypotheticals") were handled as alternative interpretations of (almost) the same system. *Begriffsschrift* presented the first integrated logic, with a predicate calculus *based on* a sentential calculus. It was of course his employment of individual variables to express generality (together with a device—the quantifier—for demarcating the scope of the generality thus expressed) which enabled Frege to base quantification theory upon the logic of truth functions. In turn, the crucial step for the introduction of variables of quantification was his 'replacement of the concepts *subject* and *predicate* by *argument* and *function* respectively' [*Begriffsschrift*, Preface]. Now to construe sentences as completed functional expressions is *ipso facto* to construe them as proper names, for a completed functional expression *is* a proper name: 'we obtain a name of the value of a function for an argument, if we fill the argument-places in the name of the function with the name of the argument' [*Grundgesetze*, §1]. According to the functional theory of predication, therefore, a sentence subsuming an object under a concept is a proper name of the value of that concept for that object as argument. (And if singular sentences are proper names then all sentences are proper names. For the extremity of the horizontal [*Grundgesetze*, §5] is accessible alike to singulars and to other sentences; whereupon those other

<sup>1</sup> Particularly to 'Function and Concept' and 'On Concept and Object' in Geach and Black, *op. cit.*; to *Nachgelassene Schriften*, p. 192, pp. 209f. Even if *Grundgesetze*, §29 suggests that Frege employs nothing more than *virtual* predicational quantification, his second-order logic is not in fact predicative.

sentences must equally be names, on pain of grammatical outrage to the horizontal).

I am not of course offering to defend the functional theory of predication or its corollary that sentences are names. I have no objection to Tugendhat's discarding these doctrines. (On the contrary, I share what I take it is the majority view, that, although the conception of the singular sentence as a completed functional expression is historically important, as having provoked the notion of an *open sentence* and hence provided the genesis of quantification theory, there remains, now that that innovation has established itself, no further motive for insisting that predicates form *singular terms* out of singular terms. The status of the functional theory of predication is that of a fruitful metaphor; nothing more.) What I do object to is these doctrines' being discarded *in Frege's own name*. Tugendhat's notion of TVP is advanced as a mere modification of Frege, as a mildly variant account which 'permits a better understanding of Frege's own exposition' [p. 182; cf. pp. 181, 183]. But in seeking to deny that the *Bedeutung* of a sentence is an object which it designates, Tugendhat is in fact proposing a revision of Frege so drastic, so revolutionary, as to overthrow his entire theory of predication.

In the final count, how much of Frege actually survives Tugendhat's "emendations"? First, what of Frege's ontology? Tugendhat's ground for denying that 'the *Bedeutung* of an expression is the object which the expression names' turns out always to be proposition (iii), although this vital fact is obscured by his failure to discriminate between (i), (ii) and (iii) throughout his discussion. Now it hardly wants remarking that proposition (iii) is totally alien to the canons of Frege himself: for Frege, concepts and truth values are as little to be accounted fictions as are numbers. If Tugendhat had anywhere made so bold as to opine that there is really no such thing as the *Bedeutung* either of a predicate or of a sentence, it would thenceforward have been clear that his "TVP" cannot be construed as a mere *elucidation* of Frege's '*Bedeutung*'. But Tugendhat maintains a remarkable silence about the ontic revisions implicit in his proposal. (We have furthermore seen that, given Tugendhat's explication of *Bedeutung*, it is by no means clear how—short of sheer legerdemain—we are to retain the fundamental posit that the *Bedeutung* of a singular term is its designatum—or, indeed, is anything at all.) And ontology aside, Tugendhat's reforms call for the abandonment of the functional theory of predication. I am unable to see how Tugendhat can any longer claim to be discussing 'the meaning of "*Bedeutung*" in Frege.'

## ON SEEING EVERYTHING UPSIDE DOWN

By N. G. E. HARRIS

IN 'Ontological Relativity'<sup>1</sup> W. V. Quine suggests that the riddles 'whether our neighbour may not systematically see everything upside down, or in complementary colour' (p. 201; p. 49) are alike in showing 'that things can be inscrutably switched even while carrying their properties with them' (p. 201; p. 50). The purpose of this note is to dispute whether the riddles are thus similar; colour switching would seem to be inscrutable, the inversion of what is seen is not.

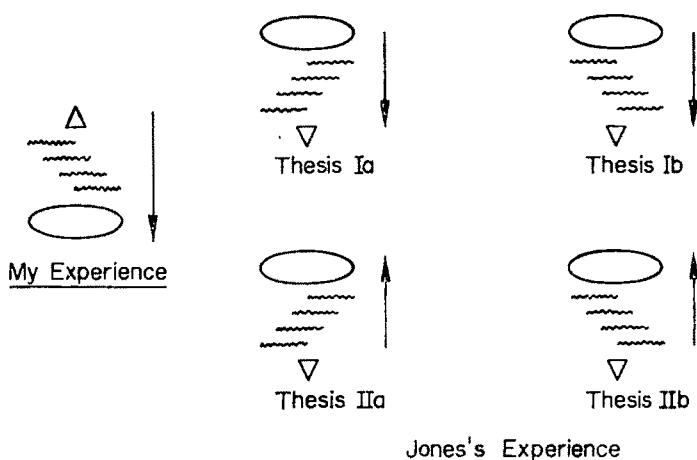
Suppose my neighbour, Jones, and I together view a mountain landscape. At the bottom of my visual field is a lake, in the middle a glacier sloping down from left to right, at the top a peak. Does Jones see the same view with everything inverted? Does he see the peak at what I should describe as the bottom of his visual field, and the lake at what I should describe as the top? Quine claims that these questions cannot be answered, yet it seems to me that we could answer them quite easily.

If I consider my own visual field I am aware of the contents of that field having position relative to felt parts of my body such as my eyelids. Whilst looking at the scene just described I should say that the part of my visual field containing the lake was nearer to my lower eyelids (perhaps we should say 'nearer to those eyelids closer to my mouth', to avoid any problems about 'lower' being ambiguous between Jones's idiolect and mine) than was the part containing the peak. Does Jones have the opposite experience? Does he have the feeling that the part of his visual field containing the peak is nearer the eyelids closer to his mouth than the rest of the scene? Since we can assume, for the sake of argument, that Jones's experiences differ from mine only in his seeing everything inverted and any differences consequential on this, we can assume that Jones also experiences relationships between the positions of objects in his visual field and his eyelids, and that he can report on them. There are two possibilities: (1) the part of his visual field containing the peak is felt by Jones to be nearer the eyelids closer to his mouth than the part containing the lake (Thesis I); (2) the part of his visual field containing the lake is felt by Jones to be nearer the eyelids closer to his mouth than the part containing the peak (Thesis II).

Furthermore, there are two kinds of inversion which might distinguish Jones's visual experience from mine. There is inversion obtained by reflection about a horizontal axis, and there is that obtained by rotation through 180°. Thus there appear to be four possible ways in

<sup>1</sup> W. V. Quine, 'Ontological Relativity', *Journal of Philosophy* Vol. LXV (1968), pp. 185-212; and in *Ontological Relativity and Other Essays* (New York and London, 1969), pp. 26-68.

which what Jones sees may be the inversion of what I see: Theses Ia and IIa (reflective inversion) and Theses Ib and IIb (rotational inversion). The diagrams below illustrate these cases with the arrows representing the direction in which the eyelids closer to the seer's mouth are felt by him to lie.



In the case of Theses Ia and Ib we might very well say that Jones saw everything upside down, but since he would describe his experience in a way that was radically different from the way I should describe mine this provides a test for the occurrence of such inversion: whereas I should describe the part of my visual field containing the lake as being *nearer to* the eyelids closer to my mouth than the part containing the peak, he would describe the part of his visual field containing the lake as being *further from* the eyelids closer to his mouth than the part containing the peak.

In the case of Theses IIa and IIb there will be no difference between the reports of Jones and myself concerning which parts of our visual fields are nearer those of our eyelids closer to our mouths, but in this case *we have no grounds at all for describing Jones's experience as seeing upside down*.

Coherent claims that something is upside down are always made (often tacitly) with respect to some frame of reference external to that of which the claim is made, as also are coherent claims that something is the right way up. In the case of such claims about the visual field of any person, the frame of reference in question must lie outside the visual field of that person. In talking of my visual field as being right way up, I must be doing so with respect to a frame of reference, and the only candidate for such a frame of reference is that provided by the tactile and kinaesthetic sensations of my body. Likewise, in talking of Jones's

visual field being inverted I must be doing so with respect to the frame of reference provided by the tactile and kinaesthetic sensations of *his* body. On the diagrams above the arrows will represent the vertical dimension of these frames of reference. But if this so, then the diagrams for Theses IIa and IIb could be, and indeed would better be, drawn thus:



That the diagrams can be so redrawn follows from the fact that what they represent is merely the positions of objects in a visual field relative to each other and to a frame of reference. Which way up the diagrams are drawn will not affect their representation of these relationships, and so the features which distinguish the two diagrams for Theses IIa, and distinguish those for Thesis IIb, are non-representational. However, since in each diagram the arrow represents the vertical dimension of the frame of reference by which inversion or its lack is judged, it is best to have it drawn in the same direction in all the diagrams to be compared.

If we now compare the diagrams immediately above with that for *My Experience* it can be seen that that for Thesis IIb is identical with that for *My Experience*, whilst that for Thesis IIa is a reflection about a vertical axis of that for *My Experience*. In neither case is there anything about the diagrams which could lead us to say that they represent cases of upside down seeing. Indeed, in the case of Thesis IIb there is just no difference at all between what Jones and I see, for the contents of Jones's visual field will be positioned relative to all other features of his experience in the same way that the contents of my visual field are positioned relative to all other features of my experience.

Although, as I have shown, the case of Thesis IIa cannot be considered as a case of upside down seeing, it may still be wondered whether the difference between Jones's experience in this case and my experience can be shown up by differences in the way each of us would describe his experience. Consider any means by which Jones and I can distinguish the left and right ears of ourselves and each other; the lack of total symmetry in the human body entails that there will always be some such means. Let us suppose, to take a simple case, that we can make the distinction in the following way: say that both of us have chilblains on what I should describe as our left ears and that our other ears are unaffected. Then I should say with reference to the contents of my visual field that the part of the glacier by the peak is *nearer to* my afflicted ear than the

part of the glacier by the lake, whereas Jones would say with reference to the contents of his visual field that the part of the glacier by the peak is *further from* his afflicted ear than the part of the glacier by the lake. So even here we have a thesis that is testable.

*University of Dundee*

## BRADY ON RECKLESSNESS

*By* WILLIAM J. WINSLADE

IN 'Indifference and Voluntariness' (ANALYSIS, Vol. 32, January, 1972, pp. 98-99) James B. Brady, in accordance with the commonly held view, defines 'acting recklessly' as 'acting [with indifference to] and awareness of a substantial and unjustifiable risk'. I have argued ('Recklessness', ANALYSIS, Vol. 30, March, 1970, pp. 135-140) that this view is mistaken. Indifference to realised risks often accompanies but is not a necessary condition for acting recklessly. A person who cares very much about the risks may act recklessly when he is ambivalent about rather than indifferent to the risks or when he fails to give due consideration to the risks.

If my view that recklessness does not imply indifference to the risks is correct, Brady's claim that '[t]he element of indifference distinguishes recklessness from acting purposely' is mistaken. Because indifference is not essential to recklessness, clearly it cannot be used as a criterion for distinguishing between acting recklessly and acting purposely. Either acting recklessly or acting purposely may, in a particular case, be accompanied by indifference. But neither implies or is implied by indifference. Suppose a war protester purposely immolates himself. Whether or not he is indifferent has no logical bearing on whether or not he acts purposely. Suppose a teenager drives his motor-cycle recklessly in his parents' presence. He may be indifferent to the risk of injury or death, but it is also possible that he is very concerned about the risks. Perhaps he drives recklessly only because he wants to frighten his parents. In any event, the presence or absence of indifference does not distinguish acting recklessly from acting purposely. (In fact, in this case we can say that the teenager purposely acts recklessly.) However, it is true that acting recklessly and acting purposely are different. Someone who acts recklessly takes a risk, but not *for the purpose* of bringing about the risked consequences. Someone who purposely (on purpose) does an act may, but need not, do so for the purpose of bringing about its consequences.

Brady makes another claim which is unclear, if not mistaken. He says that '[t]he actual awareness of the risk distinguishes recklessness from negligence; in acting negligently the agent is inadvertent to the risk'. If Brady means that in every case awareness of the risk is necessary for recklessness and that unawareness of the risk is necessary for negligence,<sup>1</sup> then he is mistaken. Although in some cases of negligence a person is unaware of a risk of which he should have been aware, there are cases in which a person is aware of a risk but negligent nevertheless. Suppose I am aware that the campfire I am building may spread to a nearby tent if I do not remove some loose dry grass next to the fire. I care about causing damage, so I take some precautions. I kick the grass a few feet away. But I fail to pay sufficient attention when kicking the grass and not all of it is removed; as a result the fire spreads, thereby causing damage to the tent. I was aware of the risk but I failed to take (enough) care to prevent the damage. Surely I was negligent.

Since awareness of risk may be present in a case of negligence, it is not an adequate criterion to distinguish all cases of negligence from all cases of recklessness. However, awareness of risk is more closely tied to recklessness than to negligence. Consider ordinary cases of recklessness, such as reckless driving or reckless use of firearms; even if recklessness does not require some awareness of risk, there is at least a strong presumption that a person who acts recklessly is to some degree aware of the risks. While awareness of risk may be present in cases of negligence, there is no presumption that it is. Indeed, in many cases a person is unaware of risks of which he should have been aware. The difference between recklessness and negligence is not captured by a single factor such as awareness of risk. A more detailed analysis of these concepts would reveal not only that both concepts have blurred edges, but also that they intersect in a variety of different ways.

<sup>1</sup> What is negligence? A. R. White has pointed out that negligence is a form of carelessness [*Attention* (Oxford, 1964), p. 82]. 'Carelessness' is defined by White as 'a failure to give active measure-taking attention to the risks inherent in the successful prosecution of some activity' (83). I agree with White; I would add that a person who is negligent is one who fails to measure up to an established standard of care.



## NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, R. G. Swinburne; J. H. Benson, Alec Fisher, Andrew Harrison, R. F. Holland, A. C. MacIntyre, Bernard Mayo, A. R. White, Peter Winch. This committee is responsible for appointing and advising the editor and for the general policy of the paper.

**SUBSCRIPTIONS.** The subscription to ANALYSIS, Vol. 33, is £1.50 for individuals in Great Britain, £1.75 for individuals abroad, and £2.00 for all institutions, post free. Each volume comprises six numbers, five of 32 pages and one of 48 pages, appearing within the academic year—in October, December, January, March, April and June. Orders should be sent to Basil Blackwell, 49 Broad Street, Oxford, or placed with any bookseller.

**CONTRIBUTIONS.** Articles submitted for publication should be addressed to Dr. C. J. F. Williams, 1 Fossefield Road, Midsomer Norton, Bath, BA3 4AS. Contributors are asked to note the following requirements.

Articles should **normally** not exceed 3,000 words in length. **Occasionally**, however, longer contributions can be accepted;

They must be typewritten in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

**It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:**

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgement is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope **and international reply coupons** of the requisite value, whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

The copyright of articles printed in ANALYSIS remains the property of the author, but contributors are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.

Vol. 33 No. 2

(New Series No. 152)

December 1972

---

# ANALYSIS

---

Edited by  
C. J. F. WILLIAMS



---

## CONTENTS

- |   |  |
|---|--|
| On the nature of justice in a trial             | G. E. M. ANSCOMBE<br><i>and</i> J. FELDMAN |
| Davidson on primitive actions that cause deaths | CONRAD D. JOHNSON                          |
| Senses of 'obliged'                             | EDGAR PAGE                                 |
| Meaning and implication: other thoughts         | R. J. KEARNEY                              |
| The truth-likeness of truthlikeness             | DAVID MILLER                               |
| Ifs and hooks: a defence of the orthodox view   | JOHN J. YOUNG                              |
| How not to reidentify the Parthenon             | FRANCIS W. DAUER                           |

---

BASIL BLACKWELL · ALFRED STREET · OXFORD

---

*Forty pence*

## ON THE NATURE OF JUSTICE IN A TRIAL

By G. E. M. ANSCOMBE *and* J. FELDMAN

THE Report on Evidence of the Criminal Law Revision Committee<sup>1</sup> would seem to give an opportunity of ending the absurd privilege (which exists in practice, although not in theory) of Roman Catholic priests to withhold from the police and the courts information learned in confessionals. Often this type of information might provide the only way to obtain the conviction of a guilty man; it is unfair to society if a criminal is able to escape punishment in this way. Indeed, there must be many Roman Catholic criminals who are escaping justice, and who could be caught were this practice changed. The truly innocent, of course, would have nothing to fear, for forcing these disclosures would only trap the guilty. Furthermore, the same standards of justice need not apply to the guilty as to the innocent, assuming it is known (say by the police) who are the guilty.

There are only four basic arguments against this development, and they are all easily disposed of. First, it might be claimed that a confessional system needs secrecy in order to operate. But all sorts of systems require secrecy; that requirement is in itself no reason to support secrecy. To the extent that criminal behaviour needs secrecy, the law should not step in with encouragement. Anyway, the psychological pressures of dependence that lead penitents to confess to priests are probably sufficiently great to guarantee that they would continue to confess even if they were aware that the information might be later used. This disposes of a second possible objection: that fear of revelation might prevent the making of confessions. Third, it might be claimed that the confessional arrangement should be encouraged; but surely that relationship is built largely on trust, and no confessee would have any reason to believe that his priest would break that trust by falsely reporting what was said. Fourth, it might be said that breaking of confessions would destroy the backbone of the Catholic religion. The Church ought to be built on a more solid foundation than that; it should not solely depend on the needs of some of its least valuable followers to make confessions of crimes against the law with impunity.

As to any argument that this advance would be unfair: the Criminal Law Revision Committee emphasized that the object of a criminal trial is to ascertain the facts, and that fairness must be measured by the accuracy of any result. Disclosures by priests would inevitably lead to

<sup>1</sup> There has appeared a recent Report by a committee set up by the Home Office in England to consider changes in the criminal law. This Report, and the model statute it contains, may well be the basis for legislation.

more, not less, accurate results. Some have implied that fairness might mean something more than just accurate results. Presumably they are worried about creating judicial procedures that, while accurate in some particular case at bar, would lead to unreliable police practices that would in turn allow the conviction of the innocent in other (future) cases. This would not be a relevant problem in the area of religious confessions.

Lest Roman Catholics be treated differently; since it is obvious that Catholics will continue to confess, the rest of society should also be encouraged to do so. The model statute appended to the Report contains a useful step in this direction when it would allow a judge, or jury, to take as corroboration of guilt any fact not divulged by a criminal at the time of his interrogation and later used by him at his trial, or also to take as corroboration the silence of the accused at a trial, where it should be his civic duty to aid in the search for truth. Perhaps science will soon be able to produce a harmless drug, put up in a vitamin pill, that would cause its users to reveal what they know about crimes. This drug could usefully be given on a regular basis to schoolchildren, who need vitamins anyway. (Indeed, we understand that one of our more forward-looking public schools has already made use of a truth-drug in its important search for truth about drug taking. Unfortunately the results in that case produced rather too much extraneous information.) Once again, the innocent could raise no appropriate objection, as they would have nothing to fear. Indeed, the general inroads made by the Report on the supposed right of silence deserve commendation. Now, when the perpetrator of some crime is accused, and feels confused and in need of help and advice, he can find safety in silence; not so in the future.

Furthermore, in many cases it is a reasonable inference from the known facts in a case that there really is no defence. The logic of the Report would suggest that the traditional archaic notions of rights to counsel in every case be sometimes abandoned. Once more, the purpose of a trial is to find out the truth. Often hearing arguments on both sides is a good way of finding out the truth. But sometimes the truth is so obvious that the adversary system becomes an absurd and wasteful game. It seems clear that no useful function could be served by having someone speak on behalf of the defendant when investigating his guilt, at least in cases where it is reasonably certain that the defendant is in fact guilty. (Similarly, in some clear cases, there should be no reason to waste valuable time advising the accused of the charges preferred against him; or to give the accused a right to confront witnesses against him.) Of course, advocates might still be useful (say) to ask for light sentences. Unfortunately, the Report itself fails actually to recommend the whole of this needed reform (in both pre-trial and trial setting) which would save much time and expense. Such a reform would actually help the adminis-

tration of justice, for it would leave defence counsel free to concentrate on protecting the innocent.

Another advance that seems implicit in the proposals of the Report would be the reintroduction of a little mild torture, under carefully supervised conditions, to obtain information. The Report might be thought, at first inspection, to rule this out on the ground that confessions obtained by torture are unreliable, and are therefore not to be introduced. That is true; but torture could often be used to obtain ascertainably reliable information. (For those who would find it undignified for judges to witness torturing: these methods would be used far away from courtrooms, well before any trial were held.) The correct position, which the Report itself puts forward, is that such evidence should be used where independently confirmable; for example, the model statute reaffirms the ability of the prosecution to produce physical evidence, or any other fact, about a crime, even though information leading to that evidence be discovered by inducements, threats, or oppressive treatment, presumably including torture.

Perhaps some might say that the justice of a trial involves something deeper than procedures designed to procure, and continue to procure, accurate results. According to them, even a man who in fact committed a serious crime, such as a brutal rape, ought only to be convicted in a certain (fair) way, where fairness means not just an accurate finding of all the facts. It means that the trial meet a number of other standards, not meeting which the authorities are a set of top bullies who, for the time being, are especially interested in the punishment of criminals. Therefore, in any fair criminal trial: a defence can be heard, innocence is presumed, and the onus of proof lies on the prosecution; infamous methods of obtaining evidence are not used or condoned; and no accused is treated before conviction in a way that would be an infringement of the basic rights of the innocent. These kinds of standards define what is a just trial. When they are not met, the "trial" is never a just trial; all are damaged by the perversion of the standards of justice, and even the criminal himself is wronged by being convicted by inappropriate procedures. That is, "court of justice" is not the same thing as "committee for nailing criminals". Similarly, 'convicting' means more than a group in power being satisfied on good evidence of the guilt of somebody, and using the powers of the state to clobber him. Of course, all of these standards are not divorced from the obtaining of accurate results—a picture of a judicial system that is both unfair and achieves accuracy would contain numerous lies. (Incidentally: an accurate result is something that can never really be guaranteed; at least an unwelcome reception of infamous methods can be assured.) This attitude would, of course, preclude the use of religious confessions in criminal trials, the denial of counsel, torture, and much else.

But all this would ultimately reduce to only a matter of opinion; there would be no logical way of proving what this fairness is, and the law should be based on facts rather than opinions.

*University of Cambridge*

## DAVIDSON ON PRIMITIVE ACTIONS THAT CAUSE DEATHS\*

By CONRAD D. JOHNSON

I WISH to discuss an argument put forward by Donald Davidson in an important recent paper.<sup>1</sup> The argument is both puzzling and revealing: puzzling because it is ambiguously formulated; revealing because even a favourable formulation must make use of certain background assumptions if the argument is to go through. And some of these background assumptions are illuminatingly false. James Cornman (*ibid.*, pp. 26–37) discusses what he takes to be a generalized version of this argument, and he ultimately rejects one of the premises in favour of another (weaker) version. But I suggest that Cornman's reformulation fails to uncover a more serious defect in the argument.

Davidson's argument is designed to show that, in cases where a person does *A* by doing *B* (the formulation 'does *A* by doing *B*' encouraging the thought that *A* and *B* are numerically distinct actions standing in a by-relation) there is really only one action being described in different ways. To take Davidson's own example, suppose that the queen killed the king by pouring poison in his ear. The queen thus did something (i.e., moved her hand in a certain way) which caused the king's death. It is tempting but mistaken, Davidson maintains, to think that the queen's moving her hand is an action distinct from her killing the king. I number four sentences of Davidson's argument because of their interest to me here:

- (1) The moving of her hand by the queen on that occasion was identical with her doing something that caused the death of the king.
- (2) Doing something that causes a death is identical with causing a death.
- (3) But there is no distinction to be made between causing the death of a person and killing him.
- (4) It follows that what we thought was a more attenuated event—the killing—took no more time, and did not differ from, the movement of the hand. (*Op. cit.*, p. 22.)

\* I am indebted to my colleague Ray Martin for encouragement and helpful criticism in writing this paper.

<sup>1</sup> 'Agency', in *Agent, Action and Reason*, ed. Robert Binkley, *et al.* (Toronto: University of Toronto Press, 1971), pp. 1–25.

(There is—as Davidson acknowledges—a further restriction to be made: We shall assume that we are speaking only of cases where a person is killed by doing something which causes his death.) The main difficulties for this argument arise, I think, first, from an ambiguity in the expression 'there is no distinction to be made' in (3). The strongest justifiable formalization of (3) is inadequate to make the argument go through. Secondly, since Davidson quite clearly intends this argument as a kind of schema leading the way to similar arguments, it is hardly clear that we will always have available identities like (1). This latter problem turns out to be closely tied to the former. To make all this clearer, let me offer a formalization of the argument using abbreviations according to the following scheme:

$A^2(x, y):$	$x$ is an act of $y$
$D^2(x, y):$	$x$ is a death of $y$
$C^2(x, y):$	$x$ causes $y$
$K^3(x, y, z):$	$x$ is a killing of $z$ by $y$
$q:$	the queen
$k:$	the king

(Hereafter—for reasons to become clear—I retain the superscript only for the causal predicate.) Following (fairly closely) both Davidson's analysis of action sentences and his view of the causal relation as corresponding to a predicate with two places for event-terms,<sup>1</sup>

(5) The queen does something which causes the death of the king just in case

$$(6) (\exists x) [A(x, q) \ \& \ C^2(x, (w)D(w, k))].$$

Further, since verbs of action should be thought of as containing an extra argument place (for terms corresponding to action-events) which they do not in ordinary English appear to contain,<sup>2</sup>

(7) The queen killed the king  
if and only if

$$(8) (\exists x) K(x, q, k).$$

How, then, are we to understand (1)-(4)? If we let ' $(\iota x)Mx$ ' abbreviate 'the moving of her hand by the queen on that occasion', the Davidsonian analysis of action sentences, would doubtless require (1) to be

<sup>1</sup> 'Causal Relations', *Journal of Philosophy*, LXIV, 21 (Nov. 9, 1967), pp. 691-703. And a word about notation. For the use of the inverted iota-operator, see, for example, I. M. Copi, *Symbolic Logic* (2d ed.), sec. 5.4. ' $(\exists! x)$ ' is to be read 'There is exactly one  $x$  such that'. See Mendelson, *Introduction to Mathematical Logic* (Princeton, N.J.: Van Nostrand, 1964), p. 79.

<sup>2</sup> 'The Logical Form of Action Sentences', in *The Logic of Decision and Action*, ed. Nicholas Rescher (Pittsburgh: University of Pittsburgh Press, 1967), pp. 81-95, especially, p. 92.

$$(1') (\iota x) Mx = (\iota x) [A(x, q) \& C^2(x, (\iota w) D(w, k))].$$

Since (2) and (3) contain a new causal predicate of action, we shall introduce a new *three*-place predicate corresponding to an *agent's* causing an event to occur. Thus ' $C^3(x, y, z)$ ' abbreviates ' $x$  is a causing by  $y$  of  $z$  to occur'. Then since (2) is quite clearly universally quantified,<sup>1</sup> it should probably be understood as

$$(2') (x)(y)(z) [(A(x, y) \& (\exists w)(C^2(x, w) \& D(w, z))) \equiv (\exists w)(C^3(x, y, w) \& D(w, z))].$$

Now one way of understanding (3) so as to allow the inference to the desired conclusion is as

$$(3A) (x)(y)(z) [C^3(x, y, (\iota w) D(w, z)) \equiv K(x, y, z)].$$

Safely assuming that a person can die only once, (1') (2'), and (3A) yield

$$(4') (\iota x) Mx = (\iota x) K(x, q, k),$$

which surely seems to be an acceptable formulation of the wanted conclusion (4). The trouble with (3A), however, is that it is too strong, and for more than one reason.

Moreover, given Davidson's analysis of action sentences, (3) is ambiguous, and might be understood not as an assertion that two three-place relations (one of which is complex) are coextensive, but that those relations are coextensive only up to (but not including) the event-place which each contains. In particular, it is natural to analyze the action of causing-the-death-of-a-person in such a way that  $y$  causes the death of  $z$  just in case

$$(A) (\exists x) C^3(x, y, (\iota w) D(w, z)).$$

And  $y$  kills  $z$  just in case

$$(B) (\exists x) K(x, y, z).$$

It also seems safe enough to conclude that (A) and (B) entail each other (though of course not by virtue of logical form) and hence that

$$(C) (y)(z) [(\exists x) C^3(x, y, (\iota w) D(w, z)) \equiv (\exists x) K(x, y, z)].$$

But (C) is not strong enough for Davidson's argument. Further, even (C) is vulnerable to Cornman's counter-example (given below). I shall argue, however, that even if Cornman's counter-example is put aside, (C) cannot be sufficiently strengthened for the argument to work.

To return, however, to the claim that (3A) is too strong: Cornman in his comments on Davidson's paper has already recognized one

<sup>1</sup> A direct reading of (2) does not justify extending this remark to the variable ' $z$ ', but since it is clearly needed in the argument, I strengthen (2) by doing so.



problem with half of (3A), namely whether the left constituent is sufficient for the right: Jones, who mistakenly wires a switch causing a moon rocket to explode on take-off, causes the death of the crew, but, Cornman maintains, Jones does not thereby *kill* the crew (*op. cit.*, p. 34). My argument does not depend for its success on the success of Cornman's counter-example. What is interesting, however, is Cornman's belief that the remaining half of the biconditional is all Davidson needs for his argument to go through (p. 35). It should be clear that even the corresponding half of my (3A) will not suffice, in conjunction with (1')-(2'), to derive (4'). Nor does it seem that Cornman's formulation of what he takes to be Davidson's general argument is nearly strong enough. Thus Cornman takes the conclusion of Davidson's general argument to be that all killings of someone are bodily movements, i.e. (I take it):

$$(D) (x)(y)[(\exists z)K(x, y, z) \supset B(x, y)].$$

(D) would only give the conclusion that the killing of the king by the queen is (predicatively) a bodily movement of the queen. But, generally, if a killing is *y*'s deed, there will not always be a unique event *x* satisfying the description 'bodily movement of *y*'.

But for argument's sake let us put aside the Cornman objection in order to see other ways in which (3A) is too strong. Notice that (3A) entails both

$$(9) (y)(z)[(\exists!x)K(x, y, z) \supset (\exists!x)C^3(x, y, (w)D(w, z))]$$

and

$$(10) (y)(z)[(\exists!x)C^3(x, y, (w)D(w, z)) \supset (\exists!x)K(x, y, z)].$$

If (3A) is a correct reading of (3)—and there is good reason to believe it is—Davidson has implicitly assumed (9) and (10). And even if (3A) is stronger than Davidson intends, surely half of it—that the left constituent is sufficient for the right—is wanted. And to insure the conclusion, surely (10) is wanted as well.

There are solid reasons for denying (9), however, and these point to more serious problems for the argument as a whole. An initial supply of counter-examples is provided by situations in which an individual *P*'s doing *A* causes *P* to do *B* (voluntarily); *B* then causes *C* to occur. By transitivity of the causal relation, both *A* caused *C* and *B* caused *C*. And if this is so, there is more than one act of *P* that caused *C*. Varying the example a bit from killing, suppose I agree with a student to give an automatic 'D' if I have received no paper by 4.30 p.m. on Friday. No paper arrives by then and I automatically turn in a 'D'. Here one can argue that both the acts of agreeing and of submitting the grade caused,

let us suppose, the student's grade print-out to read 'D'.<sup>1</sup> It is not difficult to alter this example so as to impugn (9) directly: *P* agrees to murder *Q* in case *Q* turns out to know too much. *Q* does indeed, and *P* thus murders *Q*. Here there is only one killing, but—by the above line of argument—more than one death-causing act of *P*.

Whatever one may think of the bearing that voluntary acts have on the transitivity of causality, there are other reasons for rejecting (9). Consider again an example so familiar in the recent literature on action-events: the stabbing to death of Caesar. Add a bit to Shakespeare's stage directions so that Casca stabs Caesar on two occasions, and imagine that each stabbing is sufficient for the death. These stabbings are numerically distinct actions of the same individual: They occupy different stretches of time and take place in different spatial regions. What Davidson has to say about such a case of causal over-determination is worth quoting because—I suggest—it is in large part false (albeit illuminatingly so):

... Brutus's stab was sufficient for Caesar's death. Yet it was not the cause of Caesar's death, for Caesar's death was the death of a man with more wounds than Brutus inflicted, and such a death could not have been caused by an event that was *P* ('*P*' was chosen to apply only to stabbings administered by a single hand). The trouble here is not that the description of the cause is partial, but that the event described was literally (spatio-temporally) only part of the cause. ('Causal Relations', p. 699.)

These remarks are extremely puzzling. First, if we are looking for *the* event causing Caesar's death, of *course* we will be required to take the whole event (or perhaps, following Goodman's calculus of individuals, the *sum* of particular events)—otherwise there will be no *single* event satisfying the description 'caused the death of Caesar'. But why does Davidson want to say that the separate stabs (of Brutus, Casca, and the others) did not *each* cause the death? Presumably because *this particular* death was the death of a man with more wounds, etc., (fill in as many details as you wish). But not even Davidson wants to say that the truth of a causal statement about two *particular* events depends upon the way those events are described. The way in which the events are described matters rather to the linking of singular with general causal statements, and for this it is important that (roughly, allowing quantification over descriptions) *there are* true descriptions of those events such that, under those descriptions, those events instantiate a law ('Causal Relations', pp. 699–701). In the present example, we are surely safe in supposing that each of Casca's stabbings satisfies a description under which together with Caesar's death (under an appropriate description) it instantiates a law. For *ex hypothesi*, each of the stabbings was sufficient.

<sup>1</sup> I am indebted to Ray Martin for drawing my attention to such examples. See also Joel Feinberg, 'Causing Voluntary Actions', in *Metaphysics and Explanation*, ed. W. H. Capitan and D. D. Merrill (Pittsburgh: University of Pittsburgh Press 1955), and Hart and Honoré, *Causation in the Law* (Oxford: Oxford University Press, 1959).

These considerations also count against another part of arguments like (1')-(4'). This concerns premise (1'). In this case, (1') guarantees that  $(\exists!x) [A(x,q) \ \& \ C^2(x, (w)D(w,k))]$ , and together with (2'), it is guaranteed that  $(\exists!x)C^3(x,q,(w)D(w,k))$ . But surely if this sort of argument is to work generally (i.e., is to have available premises like (1')), we need

$$(11) (y)(z)[(\exists x)C^3(x,y,(w)D(w,z)) \supset (\exists!x)C^3(x,y,(w)D(w,z))],$$

and if what I have said is correct, this is not available.

One important reason, then, why this kind of argument fails is the falsity of (11). There can often be several distinct events such that each causes  $(C^2)$  a result, and similar remarks would apply to the sibling relation  $C^3$ . In such cases, there may, however, be just one killing by one agent (for example, if Casca alone had stabbed Caesar). Because of these considerations, I think, we should also be wary of identifying a causing act with the killing in cases where there is only one causing act. Thus an interpretation of (3) entailing that  $(x)(y)(z)[C^3(x,y,(w)D(w,z)) \supset K(x,y,z)]$  seems too strong, simply because there may be many acts (of one person) each causally sufficient for the death and yet only one killing. For related (though more general) reasons, we should doubt the corresponding conditional running in the opposite direction. If a killing is sometimes numerically distinct from the relevant act causing the death, we should hesitate before endorsing the principle that all killings are death-causing acts. Thus the strongest justifiable reading of (3) may be no more than (C) supplemented, perhaps, with (10).

Whether or not a killing of  $z$  by  $y$  is identical with  $y$ 's causing  $z$ 's death seems to depend on the particular circumstances of the killing (in particular, of the act causing the death)—circumstances that are not revealed in the descriptions ' $y$  causes  $z$ 's death' and ' $y$  kills  $z$ '. Some such acts of causing a death are temporally so near the death thus caused that ordinary language and common sense would have to be strained to draw temporal boundaries fine enough to count them as distinct. Thus a certain *kind* of stabbing may have temporal boundaries indistinguishable from those of the death, and both of these would likewise be temporally indistinguishable from the killing (there being no other event that could count as the killing). In other cases there may be a considerable hiatus between the causing act and the death.

*University of Maryland*

## SENSES OF 'OBLIGED'

By EDGAR PAGE

A SECTION of Professor White's 'Meaning and Implication' (*MI*), *ANALYSIS*, October 1971, is devoted to the rebuttal of two arguments for the view that there are two senses of 'obliged'. The first of these arguments draws on the fact that

- (1) A was legally, or morally, obliged to do X

implies 'A had an obligation to do X' while

- (2) A was obliged by the weather to do X

does not. The second argument draws on the fact that (1) does not imply 'A did X' although (2) does. White attributes both arguments to H. L. A. Hart (*The Concept of Law*) and the second to one of his colleagues. Having myself (in 'On Being Obligated' (*BO*), *Mind*, April 1973) appealed to such considerations in support of the view that there are two senses of 'obliged' (and, incidentally, being one of Professor White's colleagues) I should like to defend this approach.

The main object of *MI* is to show that some philosophers have relied on a fallacious general principle in arguing for different senses of 'obliged' and various other words. I shall not discuss this principle. Firstly, because in the form in which it is stated it is so obviously fallacious that it is less than generous to suppose anyone to have used it. Secondly, because the statement of the principle is left 'deliberately wide'—for no clear purpose—and indeed it is hard to see how its inventor would make it more precise. Nor shall I consider the examples of words other than 'obliged' which, according to White, have wrongly been said to have more than one sense.

In the final paragraph of *MI*, White says 'Although a word does sometimes have two different senses and although this is sometimes the explanation for differences in implication, too easy a recourse to this ploy deflects one's attention from the real source of the differences. . . .' So he would evidently agree that, in the absence of a better explanation, the kind of difference between (1) and (2) with which we are concerned would indicate different senses of 'obliged'; but he thinks there is a better explanation. He believes that the relevant differences are due to the difference in the kind of 'obliging factor' involved, i.e. the factor by which A is said to be obliged. For example, he thinks (1) implies 'A had an obligation to do X', while (2) does not, because (1) says, or implies, that A is obliged by a rule-bound position to do X. This fits with his thesis that to have an obligation is to be obliged by a rule-bound position, as opposed to being obliged by (physical) circumstances, to do something.

The main aim of this paper is to reject White's way of accounting for the differences between (1) and (2). This will involve an examination of the position elaborated in his paper 'On being obliged to act' (*OA*), in 'The Human Agent', ed. Vesey, 1968, to which we are referred in *MI*. In *OA*, White in fact offers an account of the meaning of 'obliged' which is clearly intended to give the sense of the word in both (1) and (2); and his claim, in *MI*, that the differences in implication between (1) and (2) are due to the differences in kind of obliging factor is dependent on this account of the meaning of 'obliged'—together with his analysis of obligation which is grafted on to it. I shall argue that both his account of the meaning of 'obliged' and his analysis of obligation have to be rejected.

White claims that, in all cases, to be obliged to act is to have one's freedom restricted; that '... what we are obliged to do we have no option but to do' (*OA* p. 78). This is the central feature of his analysis of 'obliged'. Now certainly, (2) implies 'A had to do X'. But it is important to realise that this implication is dependent on the fact that (2) also implies that A had some (unspecified) objective which he knew he could not hope to achieve, in the prevailing conditions, unless he did X. In the assertion of (2) A's implicit objective is taken for granted as legitimately determining his choice of action (cf. my *BO*) and, as there is the clear implication that doing X was necessary in the circumstances for the attainment of the objective, (2) implies 'A had no option but to do X'. The force of this can be made clear by saying that (2) implies 'A had to do X otherwise he would not have achieved his implicit objective'.

An important feature of White's account is that he believes that 'there is always an "otherwise" in the offing' with 'obliged' (*OA* p. 71; cf. also Nowell Smith, *Ethics*, p. 202). This is perhaps the source of his confusion. He wrongly believes that just as (2) implies 'A had to do X otherwise he would not have achieved his objective', so (1) implies 'A had to do X otherwise he would have acted illegally'. But this last statement implies that A was concerned to avoid acting illegally, which is not implied by (1) although it may be true. By contrast (2) does imply that A's (overriding) concern was to achieve his objective. Because of this difference (2) supports the 'had to ... otherwise ...' claim whereas (1) does not; and the fact that (1) does not removes the basis for the suggestion that it implies 'A had no option but to do X'.

Although (2) implies that A had no option but to do X, we ought to be struck by the absurdity of the suggestion that (1) carries a similar implication. What sense would it make to say that a man who freely decided not to do what he knows he was legally obliged to do had no option but to do it? Clearly none. White recognises this fact and goes to some lengths to reconcile his general position with it. So it may be thought that I am misrepresenting him; in particular, that I have

attached too much importance to what he says about 'otherwise' riders and to the fact that he sometimes expresses his view in terms of the 'had no option but to' locution.

At one point, where he is clearly dealing with the obvious difficulty that one can choose, or decide, not to fulfil one's obligations, he says 'What I have a moral or legal or logical (*sic*) obligation to do is what, given moral or legal or logical conditions, I have no option but to do' (OA p. 72). This introduces a qualification to the suggestion that (1) implies that A had no option but to do X; but the question is how to understand it. One possible interpretation is this: if A is legally obliged to do X, then X is the only course open to him according to the law or from the legal point of view. This, I think, would fit with several other passages in OA. However, it runs into difficulties in connection with conflicting legal obligations, for suppose that X and Y are incompatible actions and that A is legally obliged to do X and also legally obliged to do Y. We cannot now say that X (or Y) is the only course open to A from the legal point of view. And the same problem arises in connection with conflicting moral obligations.

Another strand in White's argument is that what I am obliged to do is 'something I am unable not to do without incurring certain consequences' (OA p. 71). This suggests a way of dealing with conflicting legal obligations. If someone has conflicting legal obligations then he is legally obliged to do two incompatible actions either of which he can only not do at the cost of behaving illegally. However, we cannot carry this over to moral obligations. A man who fails to fulfil one of two conflicting moral obligations because it is clear to him that he ought to fulfil the other does not fail in the one obligation *at the cost of behaving in a morally wrong way*. Furthermore, there would seem to be a problem, on this account, of how to distinguish between the force of 'ought' and 'obliged' in moral contexts. If X is what I ought to do, from the moral point of view, surely I can only not do X at the cost of behaving in a morally wrong way.

There is of course some truth in the claim that if I am obliged to do X I cannot not do X without incurring certain consequences. Being obliged by law to remove my car from a restricted parking area, I can only not move it at the cost of incurring a fine; but the fact that, nonetheless, I decide not to move it indicates that I am indifferent to the fine (the consequences) or have no over-riding concern to avoid it. For this reason I could not claim that *I shall have to move my car otherwise I shall incur a fine*. However, a fine is the kind of consequence that we generally expect people to avoid if they can; and the importance of this should not be missed. The claim that one cannot not do what one is obliged to do without incurring certain consequences implies that the relevant consequences are of a kind that we can reasonably expect people

to try to avoid incurring. This points up a connection between the sense of 'obliged' in (2) and its sense in the kind of case just considered. However, it does not justify the claim that there is always an 'otherwise' in the offing; and it is at least an open question whether a man who is *morally* obliged to do something can only not do it at the cost of incurring certain consequences which he is concerned to avoid or could reasonably be expected to try to avoid.

Unless we could decide not to fulfil our obligations (e.g. not to do what we are legally, or morally, obliged to do) it would make no sense to speak of having or deciding between conflicting obligations. White recognises this and its relevance to the sense of (1). But in his desire to show that what can be said about the sense of 'obliged' in one context can be said about it in all contexts, he blunders into saying 'In morals, as in nature, one factor may oblige me to do one thing and another factor may oblige me to do the opposite' (*OA* p. 80). But in *nature* this is plainly not so. His example 'The steepness of a hill may oblige me to change gear while the risk of damaging a faulty gearbox may oblige me to stay in the same gear' is simply contradictory. In the past tense (ignoring the puzzling occurrences of the word 'may') it would imply 'I changed gear and stayed in the same gear'. Here he is illicitly assimilating the sense of 'obliged' in (2) to the sense it has in (1), although for the most part his account of its sense is modelled on the sense it has in (2).

White's account of *obligation* is built on to his analysis of the meaning of 'obliged'. Consequently, as the latter is unsatisfactory so is the former. However, what he has to say about obligations cannot be put right by tinkering with his view of the meaning of 'obliged'. He says that having an obligation is being obliged by a rule-bound position. I shall make two further objections to this.

The first is that we can be obliged by rule-bound positions without having corresponding obligations. In playing chess I may find myself obliged by the rules of the game to move out of check, or to make some other move, but it would be odd to say that I had an obligation to make the move. A student may be obliged by the regulations to take an ancillary course, but it would be odd to say that he had an obligation to do so. There are many cases of being obliged by rules, or by a rule-bound position, where 'obligation' has no place. Part of the explanation of this is that 'obligation' is more of a *moral* concept, even in contexts where we contrast moral and non-moral obligations (e.g. legal and moral obligations), than is allowed for on White's account; but I shall not develop this. More relevant to the present discussion is the fact that in the above examples, concerning chess and student regulations, 'obliged' functions as it does in (2), rather than as it does in (1). 'A was obliged by the rules of chess to move out of check' implies 'A moved out of check'; and 'A was obliged by the regulations to take an ancillary course'

implies 'A took an ancillary course'. (If A resigned from the game, or gave up his studies, we should say that had he continued he *would have been* obliged to move out of check, or to take an ancillary course.) They respectively imply 'A had to move out of check otherwise he would not have been playing according to the rules' and 'A had to take an ancillary course otherwise he could not have achieved his objective (say) of getting a degree'. And there are the corresponding implications that A had the objective of playing chess according to the rules, on the one hand, and that A had the objective of getting a degree, on the other. This runs contrary to White's general account of the kinds of implication that arise as depending on the kind of obliging factor involved. It runs counter to his explanation of the differences in implication between (1) and (2).

The second objection to White's account of obligation is that, in so far as we are concerned with moral obligations, it gives too prominent a place to rules. Where we are concerned with legal obligations identification of rules of law *as laws* is logically prior to a claim that someone is legally obliged, or obliged by law, to do something. But suppose I encourage a promising student to undertake research in philosophy. I may subsequently come to have a moral obligation to the student to do what I can to help him to get a job. Now it may be reasonable to recast this as my being obliged by my moral position to help the student to find a job. However, it is questionable, to say the least, whether my moral position can be defined by reference to a set of rules that can be identified as *moral* rules independently of the judgement that in this situation I have a moral obligation to the student. I am not, of course, thinking of a case where I am morally obliged by my position as the student's supervisor, say; in such a case we could no doubt identify rules of a sort, laying down what I should do. In the case I have in mind I may have no formal relation with the student. Perhaps a general principle could be produced to cover this case, laying down what must be done in such and such a kind of situation (cf. *OA* p. 67), but it would need to be recognised as a moral principle. In the end it would be recognised as such through recognition that cases falling under it were cases in which one person has a moral obligation to another. In this kind of situation the rules do not 'impose obligations' (*OA* p. 67). It is rather that our judgements—that people have, or would have, moral obligations in such and such situations—give rise to such rules as we have. This issue is, of course, connected with the wider question concerning the place of rules in morality generally, which I shall not discuss.<sup>1</sup>

<sup>1</sup> I am grateful to my colleagues Paul Gilbert and David Walker for discussion of a draft of this paper.



## MEANING AND IMPLICATION: OTHER THOUGHTS

By R. J. KEARNEY

IN his 'Meaning and Implication' (ANALYSIS, Oct. 1971) Alan R. White discusses a principle, 'If  $pr$  implies  $q$ , while  $ps$  does not imply  $q$ , then the sense of 'p' in the former is different from its sense in the latter'. He wishes to show 'that (a) this principle has been commonly used by philosophers; that (b) the principle is fallacious; and (c) the problems it has been called on to solve can be solved by other means'. His method is to discuss four examples of what he alleges to be employment of this principle, with argument that in each case reliance on the principle is not justified. For a large part his argument consists in the appeal to analogy with other cases in which employment of the principle might more clearly be seen to be fallacious, on the grounds that in those cases the principle would be warrant for conclusions that are untenable.

Since White does not 'wish to suggest that any philosopher would subscribe in general to this principle', and indeed, as he observes, 'in many instances its fallaciousness is glaringly obvious', the interest of the discussion presumably lies in the actual examples discussed in support of the contention that nevertheless the principle has been commonly used by philosophers and in the method used to make manifest the fallaciousness of the principle in those particular cases. The purpose of the present comments is to question, in two respects, the adequacy of White's discussion of those examples. It will be argued that (1) in three of the cases cited by White as examples of the allegedly common use of the principle it can reasonably be denied that they are indeed examples of employment of that principle; and that (2) in some of the cases presented as instances of analogous reasoning leading to untenable conclusions either (a) there is not the required parity of reasoning or (b) it is not at all beyond question that the conclusions are untenable.

### 1. THE EXAMPLES CITED

The first example cited by White is attributed to H. L. A. Hart, *The Concept of Law* (1961), pp. 79, sqq., in which Hart is construed as using the questioned principle in order to argue that 'there are two senses of "oblige"'. In a careful reading of the pages referred to I have failed to discover quite this argument, but rather an argument merely for the conclusion that 'There is a difference . . . between the assertion that someone *was obliged* to do something and the assertion that he *had an obligation* to do it' (Hart, op. cit., p. 80). Quite unlike the principle White was discussing, the principle of Hart's argument appears to be something

like: If  $p$  does not imply  $q$ , then the sense of 'p' is different from the sense of 'q'.

As a second example White cites Hart and Honoré (*Causation in the Law*, (1959), p. 61) as arguing that 'there are at least two senses of "responsible"', since *He was (causally) responsible for the damage* implies *He did the damage*, whereas *He was (legally) responsible for the damage* does not imply *He did the damage*'. As one might infer from White's use of parenthesis, Hart and Honoré are not considering the use of the word 'responsible' in cases in which the qualificatory words 'causally' and 'legally' are actually used, but rather its use in the possibly ambiguous cases in which no such qualifications are expressed. That is, their argument is this: It can be the case that *He was responsible for the damage* implies *He did the damage*, and also it can be the case that *He was responsible for the damage* does not imply *He did the damage*, therefore there are two senses of 'responsible'.

It is not at all clear that this argument employs the principle White is concerned to expose. It seems reasonable to interpret the argument as employing a more restricted principle, namely: If it can be the case that  $ps$  implies  $q$ , and also it can be the case that  $ps$  does not imply  $q$ , and if in both cases 's' has the same sense (and reference), then there are at least two senses of 'p'. An argument employing White's principle requires consideration of three different sentences  $pr$ ,  $ps$  and  $q$  (or at least the principle is not restricted to the case in which  $r$  and  $s$  are identical); but this argument makes comparison of only two,  $ps$  and  $q$ .

Of the two other examples discussed by White that which he cites from Price is also open to query. The argument alleged as employing the fallacious principle is this: 'Since believing in one's doctor implies that one's doctor exists, whereas believing in fairies does not imply that fairies exist, therefore there are two senses of "believe in"'. In this instance the query is not whether there are too few sentences under consideration for it to be an application of White's principle, but whether there are too many. Here there are not three sentences— $ps$ ,  $pr$  and  $q$ —but four: 'He believes in his doctor', 'His doctor exists', 'He believes in fairies', 'Fairies exist'. The use of White's principle would produce rather the argument: *He believes in his doctor* implies *His doctor exists*, while *He believes in fairies* does not imply *His doctor exists*, so the sense of 'He believes' in the former is different from its sense in the latter. Manifestly this is not Price's argument.

Implicit in White's presentation of Price's argument might be perhaps the assumption that an equivalent restatement of it would be: Since believing in one's doctor implies that what one believes in exists, whereas believing in fairies does not imply that what one believes in exists, therefore there are two senses of 'believes in'. Here of course there are the required three sentences corresponding to  $pr$ ,  $ps$  and  $q$ ,

namely: 'He believes in his doctor', 'He believes in fairies' and 'What he believes in exists'. Unfortunately, however, even formulated in this way it is not at all evident that Price has used any such argument. In a careful reading of the twenty-eight pages of argumentation to which White gives reference I have failed to recognize an argument of this simplicity. If a brief formulation of Price's argument is possible at all, it is certainly not this.

In one or more of these three examples it might be indeed that the reasoning is fallacious; but, if so, the fallacy is not that of employing the principle enunciated by White.

## 2. PARITY AND FALLACY

### (a) *Is there Parity?*

To the extent that the arguments White cites do not use the principle he wishes to discuss, to that extent it cannot be that there is parity of reasoning with any argument that does use that principle. In at least one instance there is clearly not the parity of reasoning that White claims. The argument attributed to Hart and Honoré, concerning the sense or senses of 'responsible', is said to be 'no better than the analogous, but clearly fallacious, argument that since *He was vicariously responsible for serving drink to persons under 18* implies that he did not himself serve the drinks, whereas *He was non-vicariously responsible for serving drink to persons under 18* does imply that he served the drinks, therefore, there are two senses of "responsible".' As has been shown above, the argument of Hart and Honoré is one in which a single sentence 'He was responsible for the damage' could have ambiguously the sense of either of the two different sentences 'He was (causally) responsible for the damage' and 'He was (legally) responsible for the damage'. For there to be the required parity of reasoning, therefore, it must be that the single sentence 'He was responsible for serving drink to persons under 18' could have the sense of either *He was vicariously responsible for serving drink to persons under 18* or *He was non-vicariously responsible for serving drink to persons under 18*. White does not suggest that this could be the case. To say that someone was responsible for serving drink to persons under 18 is not ambiguous, but undetermined, as to whether he was vicariously or non-vicariously responsible. *He was responsible for serving drink to persons under 18* is no more ambiguous in this respect than *It is blue* is ambiguous between *It is dark blue* and *It is light blue*. (*He was responsible for serving drink to persons under 18* might be ambiguous between *He was non-vicariously responsible for serving drink to persons under 18* and *He was at least vicariously responsible for serving drink to persons under 18*; but this ambiguity would suit White's purpose less well, since it is not clear that here 'responsible' does not have two senses.)

Others of White's examples might reasonably be interpreted in such a way as to justify the claim of parity of reasoning, but, as has been seen, only at the cost of ceasing to interpret them as employing the principle that White had undertaken to discuss.

(b) *Is there fallacy?*

Whether there is the required parity of reasoning or not, and whatever the precise nature of the reasoning, in some of his examples White's insistence that the reasoning leads to untenable conclusions is somewhat stronger than seems warranted. In objecting to reasoning that would lead one 'to invoke two senses of, e.g. "have", "require", "demand"' (in such pairs of sentences as 'Owing to the storm everyone had to abandon ship' and 'According to the regulations everyone had to have two sponsors') White asserts that 'Such an extension of ambiguity is fantastic'. In the case of 'have', for example, my own intuitive reaction is that it is far from fantastic to suggest that the word might have more than one sense. This word occurs in ambiguous sentences (e.g. 'Whenever there was a storm, they had to abandon ship' could have the sense of either *There was a regulation requiring that the ship be abandoned whenever there was a storm*, which does not imply that in fact the ship was actually abandoned, or *The effect of a storm was to make it impossible to do other than abandon ship*, which does imply that the ship was actually abandoned). While White *may* be correct in denying that the ambiguity of such sentences is due to 'have' being used with two senses, that he is correct is not so obvious as to require no supporting argument.

*University of Tasmania*

## THE TRUTH-LIKENESS OF TRUTHLIKENESS

By DAVID MILLER

'WHAT is truthlikeness?' said G. S. Robinson, and would not stay for an answer. Or not, it seems, for very long.

(i) In a recent note<sup>1</sup> Mr. Robinson fabricates such a fantastical misrepresentation of Popper's theory of verisimilitude that the censure he attaches to it, that Popper 'wrongheadedly [seeks] . . . formal and mechanical definitions of such notions as progress or scientific advance' (p. 196), is almost a matter of adding insult to imagery. Mr. Robinson accuses Popper of 'concealing [a suspicion of subjectivity] . . . under a

<sup>1</sup> G. S. Robinson, 'Popper's Verisimilitude', *ANALYSIS* 31, 6, June 1971, pp. 194-196.

cloak of formal procedures, and [of allowing] . . . intuitive assessment to masquerade as the outcome of logical and mathematical calculation' (pp. 194f.). And he goes on, even more sternly: 'This sort of situation is all too familiar in these days when so much work . . . consists in the mistaken application of mathematical and logical formalisms to inappropriate materials, giving a wholly false sense of accuracy and objectivity.'

There is of course no suggestion of swindling. But nor is there any acknowledgment of Popper's lifelong opposition to the method of overformalization, an opposition in fact registered quite unambiguously in the very paper to which Robinson refers.<sup>1</sup> I need hardly add that Popper's resistance to symbolmindedness is not a resistance to mathematics and logic themselves, or even to formalization. Indeed, "formal methods" and even "full" formalization are—in relative consistency proofs, for example—in some cases essential steps in the solution of problems. But no problem is solved by formalization alone, and we must not mistake the achievement of carrying out the formalization (which may be a considerable achievement) for the achievement of solving the problem on whose behalf the formalization was carried out.<sup>2</sup> Robinson, however, appears not to distinguish in this way between profitable applications of mathematics and others; for him, it seems, mathematical methods, if brought to bear on 'inappropriate materials' (like the growth of science), are 'useless', regardless of the problem at hand. (He gives no reason.) In my opinion, this has discouraged him from considering seriously—or even recognising—the problem at Popper's hand.

(ii) This problem, stated clearly on pp. 231f. of *C. & R.*, is simply that 'many people have been moved to say that even though we do not know . . . how far we are from the truth, we can, and often do, *approach more and more closely to the truth*. . . . Yet . . . [there seems to be] a vague and highly metaphysical notion [of truth] . . . involved here. . . . Is it not dangerously misleading to talk as if . . . we can sensibly say of two

<sup>1</sup> 'Among the real dangers to the progress of science [Popper writes] is . . . a misplaced faith in formalization and precision. . . .' See *Conjectures and Refutations* [*C. & R.*], Routledge and Kegan Paul, London, 1963 and later editions, p. 216. Other passages in the same spirit are to be found in *C. & R.*, pp. 221f., 273; in *The Logic of Scientific Discovery*, Hutchinson, London, 1959 and later editions, pp. 20–22, 378, 394; and, particularly, in Chapter 11, section II, of *The Open Society and Its Enemies*, Routledge and Kegan Paul, London, 1945 and later editions.

<sup>2</sup>As an example of an unrewarding feat of formalization I suggest Carnap's 1950 theory of induction. Shorn of its formal paraphernalia this theory contracts to little beyond the well-worn doctrine that degree of confirmation can be measured by probability. It gives no indication of how a symmetry (that is, indifference) principle can be consistently operated in a sensible language, and thus it solves no real problem. At the other extreme, Tarski's theory of truth not only solves the important formal problem of recursively defining truth for certain interesting formal languages. It also points to a solution, for natural (informal) languages, of all the difficulties that have beset the commonsense theory of truth as correspondence with the facts. (I thus disagree with those philosophers who represent the recursiveness of the truth definition as the main philosophical achievement of Tarski's theory.)

theories—say an earlier theory  $t_1$  and a later theory  $t_2$ . . . [—] that  $t_2$  has superseded  $t_1$ , or progressed beyond  $t_1$ , by approaching more closely to the truth than  $t_1$ ?' The problem is *not* 'How do you know that the theory  $t_2$  has a higher degree of verisimilitude than the theory  $t_1$ ?' but rather 'What do you intend to say if you say that the theory  $t_2$  has a higher degree of verisimilitude than the theory  $t_1$ ?' (C. & R., p. 234).

Thus the rehabilitation (as Popper has called it elsewhere)<sup>1</sup> of the idea of verisimilitude is at stake; nothing more. This is an enterprise rather akin to finding consistency proofs, and for this reason it is scarcely surprising if a few technicalities turn out to be helpful. But the technicalities in Popper's definition do not make, nor are intended to make, the objectivity. Truthlikeness as Popper defines it is objective because truth is; truthlikeness is very like truth.

I should say at once, however, that although Popper's solution is, in my opinion, objective, it is also objectionable. I shall sketch my reasons in section (vi) below.

(iii) Robinson undertakes to show that Popper is wrong in three ways in believing of 'his conception of verisimilitude . . . (1) that it is objective; (2) that it is applicable to scientific theories; and (3) that it is consonant with his [Popper's] general methodological views' (p. 194). He starts with two arguments against (2). The first (p. 195) simply misunderstands Popper's intention; the second (pp. 195f.) gets his intentions roughly right, but misstates his definition. Having thus argued that Popper's 'conception of verisimilitude' is not 'applicable', Robinson obligingly replaces it by an alternative 'applicable' conception; and this, he has no difficulty in showing (p. 196), satisfies neither (1) nor (3).

(iv) These arguments can be briefly disposed of. For Popper does not intend verisimilitude to be 'applicable' in any practical sense. (Consequently, scientists are not expected to make use of it; contrast Robinson, p. 195, lines 5–9, and p. 196, lines 15–23, with C. & R., p. 235, lines 1–9.) Likewise, when he assumes that 'the content and the truth-content . . . are in principle *measurable*' (C. & R., pp. 233f.; Robinson, p. 195), 'in principle' means what it says. Thus Robinson's accusations of impracticability are beside the point; and we should indeed 'stay firmly in a . . . theoretical realm where consequences are simply true or false without further ado' (p. 195). Robinson considers this. But because he imagines—for reasons that escape me—that Popper recommends cardinality as a measure of content ('the verisimili-

<sup>1</sup> See for example pp. 25f. and 29 of 'A Realist View of Logic, Physics, and History', pp. 1–30 of Wolfgang Yourgrau and Allen D. Breck, editors, *Physics, Logic, and History*, Plenum Press, New York and London, 1970. This paper is now reprinted as Chapter 8 of Popper's *Objective Knowledge*, O.U.P., 1972; see especially pp. 314–317.

tude of a . . . theory consists in the excess of the number of true logical consequences of that . . . theory over the number of false logical consequences'—p. 194), he does not consider it for long. Instead he offers as a last hope for Popper a 'practical' (or 'epistemic') definition of verisimilitude, which we need not discuss. Thus Robinson's whole paper is based on a sequence of misreadings and misunderstandings.

(v) Nor are the misunderstandings really understandable misunderstandings. (And to treat them seriously is not really defensible; yet such floaters are in the air these days.) For without even leaving p. 234 of *C. & R.* we can read that verisimilitude 'is *not an epistemological or an epistemic idea*—no more than truth or content [is]. . . . Accordingly . . . [the answer to] the question "How do you know that the theory  $t_2$  has a higher degree of verisimilitude than the theory  $t_1$ ?" [is] . . . "I do *not* know—I only guess. But I can examine my guess critically . . ."'. There is no more a 'criterion' for increase in verisimilitude than there is one for truth. But this hardly endangers the objectivity of verisimilitude, which 'has the same objective character . . . as . . . absolute *truth*' (*C. & R.*, p. 234). And since Robinson offers no criticism of the preceding sections of Popper's paper (pp. 223–228), where the objectivity of truth is thoroughly argued, I do not know what 'element of judgment' Robinson ('truth . . . may or may not be "objective"'—p. 194) suspects may be lurking there.

As far as the measure of content goes, it is obvious why cardinality is not used. For *all consequence classes have  $\aleph_0$  members*. Thus all truth contents have  $\aleph_0$  members. Falsity contents can of course be empty, but otherwise they too will have  $\aleph_0$  members. (Robinson seems a bit at sea here, to judge by the first few lines of p. 196.) The natural first move is to factor out by interderivability—that is, to identify sentences that are logically equivalent—but it gets us virtually nowhere. To be sure, finite truth contents become possible in principle; though of no interest in practice. Falsity contents, however, remain empty or countable, unless the set  $T$  of all true sentences in the language under consideration is finitely axiomatizable.<sup>1</sup> Thus only the minutest discrimination can be achieved by cardinality considerations; less, even,

<sup>1</sup> In this context the assumption of  $T$ 's unaxiomatizability has been made by Popper in 'A Theorem on Truth-Content', pp. 343–353 of Paul K. Feyerabend and Grover Maxwell, editors, *Mind, Matter and Method, Essays in Philosophy and Science in Honor of Herbert Feigl*, University of Minnesota Press, Minneapolis, 1966; see especially p. 349. It is a ready consequence of Theorem 1 of this paper that if  $T$  is unaxiomatizable then the truth content of any false theory is also unaxiomatizable. *A fortiori*, a false theory cannot have a finite truth content; nor a finite falsity content (consider what would be the complement of the disjunction of all the elements of such a finite falsity content). And a true theory can have a finite content only if it is the conjunction of finitely many irreducible theories (*op. cit.*, p. 348, and § 4 of Alfred Tarski, 'Foundations of the Calculus of Systems', Chapter XII of *Logic, Semantics, Metamathematics*, O.U.P., 1956); in other words, uselessly weak from the point of view of empirical science.

than what is available from a comparison of contents. Somewhat more sophistication is therefore desirable. And, indeed, when (still on p. 234 of *C. & R.*) Popper mentions 'measure', only a brief glance at the section of the *Addenda* there indicated is needed to show that he uses it in the sense of 'probability measure'. And to this Robinson's criticisms cannot immediately be said to pertain.

(vi) For indeed there is no end to the possible values that verisimilitude, defined by

$$Vs(a) = \frac{p(a, a_T) - p(a_T)}{p(a, a_T) + p(a_T)},$$

may obtain ( $a_T$  is the truth content of  $a$ ;  $p$  is a probability measure; see *C. & R.*, p. 397). Yet if we restrict our attention to universal theories  $a$  (for which  $p(a)=0$ ) we will, I think, have to set  $p(a_T)=0$  and  $p(a, a_T)=1$ , irrespective of whether  $a$  is true or false. So Robinson's strictures seem to get home after all.

Why must  $p(a_T)=0$  and  $p(a, a_T)=1$ ? Here  $a_T$ , the truth content of  $a$ , is just the product, in the lattice of deductive systems, of the systems  $a$  and  $T$ .  $T$ , as I have mentioned, is by virtue of Gödel's theorem—perhaps for other reasons too—to be regarded as unaxiomatizable. If we identify models which are elementarily equivalent, it follows that  $T$ —which is complete—has only one model. On the other hand  $a$ , if axiomatizable (even only recursively), is satisfied in infinitely many models (for as Gödel showed, any consistent axiomatizable theory containing Peano arithmetic has continuum many consistent complete extensions). Thus  $a_T$ , whose models are just the models of  $a$  together with those of  $T$ , has more or less the same models as  $a$ , whether  $a$  be true or false. Following Łos, Kemeny, and others in taking  $p$  as a measure over the set of models we can see that, unless it is scandalously biased in favour of  $T$ , we are bound to get  $p(a_T)=0$ ,  $p(a, a_T)=1$ .

Is this a serious criticism, or does it merely reflect the 'inapplicability' of verisimilitude? Largely, though not entirely, the latter, in my opinion. For just as we can often genuinely distinguish two theories  $a$  and  $b$  (each zero in probability) by contrasting  $p(a, b)$  and  $p(b, a)$ , so here we can define relative verisimilitudes in an obvious way; and a comparison of  $Vs(a, b_T)$  and  $Vs(b, a_T)$  will often indicate which of  $a, b$  is closer to the truth. For example, if  $b \vdash a$  then, as is easily checked,

$$\begin{aligned} Vs(a, b_T) &= \frac{p(a, b_T) - 1}{p(a, b_T) + 1} \\ &\approx \frac{p(a, b) - 1}{p(a, b) + 1} = 0; \end{aligned}$$



whilst

$$\begin{aligned} V_s(b, a_T) &= \frac{p(b, b_T) - p(b_T, a_T)}{p(b, b_T) + p(b_T, a_T)} \\ &\approx \frac{1 - p(b, a)}{1 + p(b, a)} \end{aligned}$$

which only equals  $V_s(a, b_T)$  if  $b$  and  $a$  are almost equivalent. Thus some sort of discrimination is achieved.

What is a good deal more unfortunate is the fact that the falsity content of any theory  $a$ , true or false, defined by  $1 - p(a, a_T)$ , is equal to zero; and this result is not to be changed by relativizing the probability measure. But if the falsity content is zero in every case, then comparison of verisimilitudes reduces to comparison of truth contents; that is, to comparison of contents. Whatever Robinson's remarks might seem to encourage, this conclusion, which suggests that we can know *a priori* that we are approaching the truth, is quite unacceptable. Do we actually have to draw it?

In my opinion we do not. We can after all, without supposing it realistically to record  $T$ 's content, allow a penchant for truth to creep into our probability distribution, so that  $p(T) > 0$ . It then turns out that, whatever value  $p(T)$  has, the falsity content of a false theory increases with its content. Since if  $b$  is false and  $b$  entails  $a$  unilaterally the false consequences of  $b$  will strictly include those of  $a$  ( $b$  is in one, though not the other) this seems a not quite implausible result. Thus to some extent it does seem possible to overcome the more distressing consequences of infinitude in this area; and to that extent Popper's theory of verisimilitude can be accorded some degree of qualitative validity.<sup>1</sup> Regrettably, once we try to replace the present (quantitative) theory by a qualitative one we are led to difficulties substantially more intractable than those considered here. This topic, however, deserves a note on its own.

<sup>1</sup> A similar attitude, though in this case an attitude to so-called probability logic, is expressed on p. 138 of Zoltan Domotor, *Probabilistic Relational Structures and Their Applications*, Technical Report No. 144, May 14, 1969, Institute for Mathematical Studies in the Social Sciences, Stanford.

*University of Warwick*

## IFS AND HOOKS: A DEFENCE OF THE ORTHODOX VIEW

By JOHN J. YOUNG

1. THE view that "factual" conditionals of the form 'if  $p$  then  $q$ ' should not be read as ' $p \supset q$ ' has become 'something of a philosophical orthodoxy', according to Mr. Michael Clark (ANALYSIS 32.2). Furthermore, he believes that the arguments put forth in support of the orthodox view by Professor Strawson in *An Introduction to Logic* (London, 1952) and David Mitchell in *An Introduction to Logic* (London, 1962) are bad ones. I shall argue that orthodoxy remains fundamentally unshaken by Clark's attack.

2. The central tenet of the orthodox position which Clark criticizes is this:

... factual conditionals of the form 'if  $p$  then  $q$ ' are not typically truth functional in meaning (p. 33).

Unfortunately Clark does not explicitly identify the factual conditional, although he seems to hold that the 'factual use' of 'if . . . then. . .' sentences corresponds to Strawson's 'standard or primary use'. Judging from Clark's comments, factual conditionals are not counterfactuals and are expressed by indicative 'if . . . then . . .' sentences. I am aware that philosophers disagree over the utility and precise significance of the term 'counterfactual', and that the indicative-subjunctive distinction has been called irrelevant to the logic of conditionals. For brevity's sake, however, I shall take Clark to be attacking the following principle, which should be acceptable to those of an orthodox stripe even if expressed in a way not entirely to their liking:

(P<sub>1</sub>) Non-counterfactual conditionals asserted or expressed by using indicative sentences of the form 'if  $p$  then  $q$ ' are not all (or, perhaps, are not typically) truth-functional in nature.

3. Almost all who hold (P<sub>1</sub>) also accept the following principle

(P<sub>2</sub>) A conditional of the form 'if  $p$  then  $q$ ' entails the corresponding material conditional ' $p \supset q$ ',

as well as

(P<sub>3</sub>) Material conditionals of the form ' $p \supset q$ ' do *not* entail that if  $p$  then  $q$ .

But Clark offers the following argument to refute (P<sub>3</sub>):

(i) If  $p, q$  entail  $r$ , then  $p$  entails that if  $q$  then  $r$ .

- (ii)  $p \supset q, p$  entail  $q$
- (iii)  $p \supset q$  entails that if  $p$  then  $q$ .

This argument, if acceptable, would indeed be powerful. For although Clark does not make the point, it would not only refute (P<sub>3</sub>), but two other claims offered on behalf of the non-hookish view:

- (P<sub>4</sub>) Although the pair of material conditionals ' $p \supset q$ ' and ' $p \supset \neg q$ ' are compatible, the pair of "factual" conditionals 'If  $p$  then  $q$ ' and 'If  $p$  then  $\neg q$ ' are *not* compatible.

and

- (P<sub>5</sub>) Although ' $p \supset p$ ' is a contingent conditional, 'If  $\neg p$  then  $p$ ' is self-contradictory.

After all, given (P<sub>2</sub>) and (iii), then 'if  $p$  then  $q$ ' and ' $p \supset q$ ' are logically equivalent. So it would follow that 'if  $p$  then  $q$ ' and 'if  $p$  then  $\neg q$ ' are indeed consistent and that 'if  $\neg p$  then  $p$ ' is a contingent claim rather than self-contradictory.

Clark's argument is formally valid, so the only course open to defenders of orthodoxy, short of heresy or apostasy, is to attack the premises. Since (ii) is true, only (i) is open to criticism. But since the truth of (P<sub>3</sub>) and its denial (iii) are the point at issue, it would be question-begging to argue, 'Given the validity of the argument and the truth of (ii), then since (P<sub>3</sub>) is true [and so (iii) is false], (i) must be false as well.'

The force of (i) seems to arise from its resemblance to the deduction theorem encountered in the study of the propositional calculus. But the deduction theorem need not commit one to premise (i). The former is sometimes stated as

If  $p, q \vdash r$ , then  $p \vdash q \supset r$ ,

a theorem about sentences in a formalized language, not about sentences in the natural language: note that the consequent contains ' $\supset$ ', not 'if'. Even if we express the theorem in terms of entailment, it only reads:

If  $p, q$  entail  $r$ , then  $p$  entails  $q \supset r$ .

We are merely *using* 'if . . . then . . .' to express the theorem, but the theorem is not about 'if . . . then . . .'. To say that (i) follows from the deduction theorem about ' $\supset$ ' would be in the eyes of the orthodox a *non sequitur* and in the context of this debate question-begging. To be sure, Mr. Clark does not say that (i) follows from the theorem about ' $\supset$ ', but neither does he offer any ground for accepting (i). Until some reasons are forthcoming, there is no need to accept this argument, nor need we regard (P<sub>3</sub>), (P<sub>4</sub>) or (P<sub>5</sub>) as refuted.

4. To refute (P<sub>4</sub>) Clark offers us a simple counterexample: a case

where sentences of the form 'if  $p$  then  $q$ ' and 'if  $p$  then  $\neg q$ ' can express consistent propositions. Consider the following:

- (C) The match won't be cancelled<sup>1</sup>
- (D) Whenever it rains on the day of a match, that match is cancelled
- (E) If it rains, the match will be cancelled
- (F) If it rains, the match won't be cancelled.

Clark then argues:

- (i) (C) and (D) are compatible
- (ii) (C) entails (F)
- (iii) (D) entails (E)
- (iv) So (C), (D) entail the conjunction (E) & (F)
- (v) (E) & (F), allegedly incompatible propositions, must be consistent, for they are entailed by the compatible propositions (D) and (C).

The objectionable step in this argument is (ii), which Clark defends by saying:

... It follows from the statement that the match won't be cancelled that it won't be cancelled whatever (or at least whatever else) in fact happens. No doubt there are events whose occurrence has not yet been made impossible, and if these *were* to happen the match would be cancelled; but it is necessary that if in fact it isn't cancelled these events aren't in fact going to occur. It is important to notice that the antecedent in (F) means 'if in fact it rains', not 'if it were to rain' (in which case the consequent would have read '... would not ...'). (p. 35.)

In opposition to Clark, it should be noted that it is wrong to hold as a general principle:

If  $p$ , then  $p$  whatever (or at least whatever else) happens.

The statement that the match won't be cancelled is simply not as strong as the claim that the match won't be cancelled no matter what (else) happens. More to the point, it is possible to believe that the match will not be cancelled (perhaps on the ground that the weather forecast is favourable) and consistently refuse to accept

(F) If it rains, the match will not be cancelled.

Yet, according to Clark, to accept that the match won't be cancelled but *reject* (F) is to be inconsistent. And this is absurd.

Clark might protest that I have confused the indicative (F) with the subjunctive

(F') If it were to rain, the match would not be cancelled,

<sup>1</sup> I use Clark's labels for examples throughout this discussion for ease of comparison.

for the antecedent of (F) is to be taken as 'if in fact it rains'. However, I am not guilty of any such confusion. The addition of 'in fact' to 'if it rains' does nothing to clarify the significance of (F). After all, one could also add 'in fact' to the antecedent of (F') as well ('if in fact it were to rain . . .'). Even if there were some difference to be noted as Clark suggests, he would still be wrong: one can consistently accept that the match will not be cancelled, yet refuse to accept 'If *in fact* it rains, the match will not be cancelled'.

5. In his willingness to give up (P<sub>4</sub>) Clark ignores the fact that two people who assert 'If it rains, the match will be cancelled' and 'If it rains, the match will not be cancelled' take themselves to be genuinely disagreeing *even if* both believe or know that it is not going to rain. If these conditionals were material conditionals, resolution of the dispute should occur by pointing out that it is not going to rain. But that is not the way such disagreements are resolved.

Here is another argument in support of (P<sub>4</sub>). Suppose that Smith asserts

(E) If it rains, the match will be cancelled,

and that (E) is a material conditional. How could one go about expressing disagreement with what Smith said? Not by asserting

(F) If it rains, the match won't be cancelled,

since *ex hypothesi* (E) and (F) are compatible. Short of saying 'I disagree with what Smith said', the only other way seems to be asserting

(E') It is not the case that if it rains the match will be cancelled.

If we represent (F) as ' $R \supset \neg M$ ' and (E') as ' $\neg(R \supset M)$ ', then (E) is equivalent to ' $\neg(R \ \& \ \neg M)$ ' and (E') is equivalent to ' $R \ \& \ \neg M$ '. So given that (E) and (F) are compatible, one must assert (or at least be committed to the truth of) ' $R \ \& \ \neg M$ ' in order to oppose (E).

But anyone who wishes to express disagreement with

(E) If it rains, the match will be cancelled,

*need not* assert (or be committed to) a claim as strong as

It will rain, and the match won't be cancelled.

Under certain circumstances, *e.g.*, when one knows or believes that it will *not* rain, one could still in good conscience assert (E'), but asserting 'It will rain, and the match won't be cancelled' would be dishonest.

Nevertheless, there is an important restriction which must be placed on (P<sub>4</sub>): it applies only to conditionals of the form 'if *p* then *q*' and 'if *p* then  $\neg q$ ' whose common antecedent is consistent or not self-contradictory.

dictory. Such a qualification is needed in the light of *reductio ad absurdum* proofs in mathematics, where we may have such "opposite" conditionals both true when the antecedent is inconsistent. Thus restricted, ( $P_4$ ) still enables us to distinguish the relevant non-truth-functional conditionals from material conditionals, since the latter, but not the former, will be true simply in virtue of a false antecedent.<sup>1</sup> A possible alternative might be to construe the 'if's of *reductio* proofs as differing from both hooks and the 'if' used to express non-material, empirical conditionals.

6. To refute ( $P_5$ ), Clark offers the following counter-example:

- (G) If he hasn't left her, he must have become tough enough to live with her
- (H) If he became tough enough to live with her, he became bold enough to leave her
- (I) If he was bold enough to leave her, he has left her

---

So, (J) if he hasn't left her, he has left her.

(Therefore, (K) he has left her.)

Clark comments, 'it must surely be granted that (G), (H), (I) entail (J), and also that (G), (H) and (I) are mutually consistent' (p. 36).

Let us assume that transitivity holds (although not all defenders of orthodoxy would agree, *e.g.*, R. Stalnaker, *Studies in Logical Theory*, edited by N. Rescher, Oxford, 1968, p. 106), and that the argument is indeed valid. Note that (H) and (I) entail 'If he became tough enough to leave her, he has left her', which by contraposition, becomes

- (G') If he hasn't left her, he has not become tough enough to live with her.

So (G) and (G') must be consistent, if (G), (H), (I) are consistent. Clark cannot give up transitivity or contraposition without thereby acknowledging that some ifs are not hooks.<sup>2</sup> So he must reject ( $P_4$ ) if his example is to work.

However, in Section 4 we showed that he failed to refute ( $P_4$ ) and in Section 5 that there are some good reasons for holding that conditionals of the form 'if  $p$  then  $q$ ' and 'if  $p$  then  $\neg q$ ' are not consistent. These reasons reflect the rules we employ for the use of 'if' in ordinary discourse, and if Clark wishes to give an account of the 'ordinary language conditional' he must take them into account. If ( $P_4$ ) is correct, Clark is wrong in saying that (G), (H), (I) are consistent, for they entail

<sup>1</sup> The need for some such restriction if ( $P_4$ ) is to be maintained was pointed out to me by Dr. Smiley.

<sup>2</sup> John Watling (ANALYSIS, 14.4) has argued that not all 'if . . . then . . .' sentences are contraposible, as has Stalnaker (*Op. cit.*, p. 107). Peter Downing suggests that some conditionals may not even have a contrapositive (*Aristotelian Society Proceedings*, Supplementary Volume XLIV, p. 43).

inconsistent propositions. In that event, the force of Clark's claim would at best be 'If (G), (H), (I) were true, (J) would be true'. But this is not enough to establish that (J) *is* true. A somewhat similar case can be made out against Clark's claim that 'if  $p$  then  $\neg p$ ' is a contingent claim.

It is worth noting that if  $(P_4)$  is true (properly qualified) we have an explanation why the formula 'if  $\neg p$  then  $p$ ' is a contradiction. I take it that a claim of the form 'if  $\neg p$  then  $\neg p$ ' is necessarily true. If this is so and  $(P_4)$  is correct, 'if  $\neg p$  then  $p$ ' must always be false.

One would rest easier with  $(P_4)$ , however, if one could show that a contradiction follows directly from a conditional of the form 'if  $\neg p$  then  $p$ ' without invoking  $(P_4)$ . Suppose that a statement of the form 'if  $\neg p$  then  $p$ ' were true. Conjoined with the necessarily true claim of the form 'if  $\neg p$  then  $\neg p$ ' we can construct the following argument:

If  $\neg p$  then  $p$ ,

If  $\neg p$  then  $\neg p$ ,

Therefore, if  $\neg p$  were true,  $p$  and  $\neg p$  would be true.

This argument is valid, for although it follows from the first premise that  $p$  is true (and so  $\neg p$  is false), it follows from, say, 'If it won't rain, it will rain' and 'If it won't rain, it won't rain' that if *per impossible* it were true that it won't rain, it would also be true that it will rain and it won't rain. But it is false that if the proposition that it won't rain *were* true then the contradiction that it will rain and it won't rain would be true. Consequently, we must infer that at least one of the premises which we *assumed* to be true is not true. The second premise, 'if  $\neg p$  then  $\neg p$ ,' however, seems to be above reproach, so it follows that 'if  $\neg p$  then  $p$ ' is false.

7. At times Mr. Clark appeals to conversational *conventions* to explain certain features of conditionals which have led some people to think that 'if'-sentences can be used to express something more than, or at least different from, what can be expressed by using the hook. Thus, conversational conventions may account for the "connection" between antecedent and consequent expressed or implied by some uses of 'if'. Again, such conventions may be responsible for the implication of uncertainty about or disbelief in the antecedent of a conditional expressed by using 'if'. Similarly with our reluctance to assert 'if  $p$  then  $q$ ' solely on the basis that  $\neg p$  or 'If ice is left in the sun it melts' merely on the ground that ice is never left in the sun.

Nevertheless Clark does not consider the possibility of invoking conversational conventions to explain the difference between our use of the subjunctive

(O) If Hitler hadn't committed suicide, Germany would have won the war,

and the indicative

(Q) If Hitler didn't commit suicide, Germany won the war.

Rather, he seems to regard (O) and (Q) as differing in "meaning", where a difference of meaning is tied to a difference in truth conditions rather than a possible difference in conversational conventions which govern the use of moods.

A case might be made out, however, for saying that the difference between (O) and (Q) lies in some such non-logical rules or implications, rather than in truth conditions. Whether this commits one to saying that the *meaning* of (O) and (Q) is the same, I shall not say. But it would serve as a ground for saying that the 'if' in (O) and the 'if' in (Q) are used the same way. Then, since Clark admits that (O) 'is false, or incorrect in some way—it certainly isn't true, anyway' (p. 37), I should suggest that the same could be said of (Q) in spite of the fact that the antecedent of (Q) is false. This line of thought seems as initially plausible as Clark's, that 'the difference in meaning between the sentences expressing (O) and (Q) could arise simply from the different tenses of the verbs' (p. 37).

Finally, appeals to conversational conventions may very well be useful, but we shall never know until such conventions are stated. Clark several times offers us promissory notes, but we shall have to await payment before we can determine their actual worth. (This holds for my own suggestion as well.)

8. Meanwhile, one cannot consistently claim to be rendering an account of 'a standard natural language conditional' (p. 37), yet ignore features governing its use. Clark's account runs counter to some important facts about our use of conditionals. In Section 5, we noted that speakers who utter indicative conditional sentences of the form 'if  $p$  then  $q$ ' and 'if  $p$  then  $\neg q$ ' in many cases take themselves to be in disagreement, not simply saying something odd or unconventional, even if both speakers know or believe that  $\neg p$ . Nor need one assert, or be committed to, the proposition that  $p$  and  $\neg q$  in order to oppose, with a clear conscience, what is asserted by uttering a sentence of the form 'if  $p$  then  $q$ '.

Also, although we can take either the proposition that  $\neg p$  or the proposition that  $q$  as a sufficient ground for the *truth* of what is expressed by ' $p \supset q$ ', in many cases they do not separately or together serve as a sufficient ground for the *truth* of what is expressed by 'if  $p$  then  $q$ '. This point is *different* from the question of the appropriateness of uttering a certain kind of sentence (except inasmuch as uttering a sentence that says what is true is preferable to uttering one that says what is false).



These considerations point to a difference in *truth conditions*, not simply to different conversational conventions or rules of helpfulness. But if what is expressed (at least on some occasions) by using a sentence of the form 'if  $p$  then  $q$ ' differs in truth conditions from what is expressed by using a sentence of the form ' $p \supset q$ ', we have on our hands two logically distinct types of proposition.

*Davidson College*

## HOW NOT TO REIDENTIFY THE PARTHENON

By FRANCIS W. DAUER

WHILE living in exile on a fashionable Mediterranean island, the Proteus brothers came across the following item in the paper:

### BOMBOS WINS CASE

OU DAMOU—The legal manoeuvrings surrounding Morion *vs.* Bombos came to an end today when the judges ruled in favour of Bombos. Several months ago Mr. Bombos brought his ship X to the Proteus brothers for a complete renewal of all its parts. It was placed in dock A. At the same time the Proteus brothers were commissioned to build a new ship for Morion in dock B. The scheming Proteus brothers hit upon the following plan: Whenever one of X's planks was replaced by a new one, the old plank was brought to dock B. In the end in dock A stood a ship Y composed of new planks and in dock B stood a ship Z composed of what were X's planks. When a fire in dock B destroyed Z, the question arose as to whose ship was reduced to ashes. Morion's lawyers argued that  $X=Z$  insofar as X's parts were identical to Z's. The judges dismissed this argument and ruled that  $X=Y$ . Mr. Smart analyzed the judges' reasoning as follows: When a plank had been removed from X and replaced by a new plank, it was the new plank that was now a part of X. The old plank *had been* a part of X but was no longer. Rather it was a part of Z which gradually came into being from hour 1 onwards. When completed Z was not composed of any of X's parts but rather of what had been X's parts . . . it was clear that X had simply changed the identity of its parts and that now its parts were identical with the parts of Y. X was not identical with Z for they had no parts in common. ('How to Re-identify the Ship of Theseus', *ANALYSIS*, April 1972, p. 148.)

The Proteus brothers immediately appreciated the significance of this ruling and it was only a matter of time before they approached Lord Elgin with the following proposition: For a "reasonable fee" they offered to complete Lord Elgin's collection by delivering to him the Parthenon of Oudamou. Returning to Oudamou, the brothers set to

work. Each night they replaced one of the ancient slabs of the Parthenon by a modern replica. Each old slab was forwarded to Lord Elgin, and in less than a decade the Parthenon of Oudamou graced his lordship's mansion.

Lord Elgin's joy turned to despair when he saw the difference the Parthenon made to the rateable value of his property. He soon joined his peers and opened his grounds to the public. His advertisements emphasized the opportunity to view the incomparable Parthenon of Oudamou. Oudamou was in an uproar. The Proteus brothers were soon arrested and charged with stealing the Parthenon.

The lawyer for the Proteus brothers argued that his clients could not be charged with stealing the Parthenon since the Parthenon was still in Oudamou and not in England. The lawyer cited *Morion vs. Bombos*. For the same reason he argued that his clients could not even be charged with stealing parts of the Parthenon. After all, its parts are to be seen on the Acropolis and not on Lord Elgin's estate.

The judges were on the verge of declaring *Morion vs. Bombos* to be irrelevant. Reidentifying ordinary ships is one thing, reidentifying a national treasure is another. They felt that where the history of the object had a decisive bearing on the object's significance the usual criteria for reidentification were inappropriate. Fortunately they were spared the task of clearly establishing this distinction between the Parthenon and Bombos' ship. At the last moment the prosecution dropped its major charge and asked only for a ruling on the several lesser charges: illegal trespass, illegal exportation of ancient marble, etc. It was later learned that the National Tourist Bureau had brought pressures upon the prosecutor's office.

The next edition of the *Guide Bleu* warned its readers of Oudamou's desperate claim that the Parthenon was where it always had been. To see the Parthenon, it recommended visiting Lord Elgin's estate.

*University of California, Santa Barbara*

## NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, R. G. Swinburne; J. H. Benson, Alec Fisher, Andrew Harrison, R. F. Holland, A. C. MacIntyre, Bernard Mayo, A. R. White, Peter Winch. This committee is responsible for appointing and advising the editor and for the general policy of the paper.

**SUBSCRIPTIONS.** The subscription to ANALYSIS, Vol. 33, is £1.50 for individuals in Great Britain, £1.75 for individuals abroad, and £2.00 for all institutions, post free. Each volume comprises six numbers, five of 32 pages and one of 48 pages, appearing within the academic year—in October, December, January, March, April and June. Orders should be sent to Basil Blackwell, 49 Broad Street, Oxford, or placed with any bookseller.

**CONTRIBUTIONS.** Articles submitted for publication should be addressed to Dr. C. J. F. Williams, 1 Fossefield Road, Midsomer Norton, Bath, BA3 4AS. Contributors are asked to note the following requirements.

Articles should normally not exceed 3,000 words in length. Occasionally, however, longer contributions can be accepted;

They must be typewritten in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgement is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope and international reply coupons of the requisite value, whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

The copyright of articles printed in ANALYSIS remains the property of the author, but contributors are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.

---

# ANALYSIS

---

Edited by  
C. J. F. WILLIAMS

---

## CONTENTS

Lost times	ROSS HARRISON
Confirmability and factual meaningfulness	R. G. SWINBURNE
Is there an Ozma-problem for time?	NORMAN SWARTZ
Are cognitive processes and structure a myth?	MICHAEL MARTIN
On promising the unwanted	W. R. CARTER
On the intrinsic wrongness of killing innocent people	LEONARD GEDDES
On the coherence of act-utilitarianism	LARS BERGSTRÖM
Reasons for acting versus reasons for believing	GARY A. WEDEKING
Martin on explanation and confirmation	BARRY GOWER
Smart on conditions of identity	LAWRENCE H. DAVIS
Russell's incomplete symbols	J. O. URMSON
Harman's non-essential property	GEORGE F. SCHUMM and JOHN G. STEVENSON

---

BASIL BLACKWELL · ALFRED STREET · OXFORD

---

*Fifty pence*

## LOST TIMES

By ROSS HARRISON

§A *Problem: Where is time's arrow?*

**S**WEPT by a wave of nostalgia for Cambridge in 1961 it seems that I can try to cure it by returning to Cambridge but not by returning to 1961. The single dimension of time, that is, seems to have a definite direction in a way that the many dimensions of space do not. This seems to be the most striking difference between space and time, and it looks as if it can be captured in the following pair of formulations:

- (1) (a) One can return to the same place.
- (b) One cannot return to the same time.

However, in this formulation 'return to' means something like 'be again at'; so the truth of (1) obviously depends on the truth of (2):

- (2) (a) One can be again at the same place.
- (b) One cannot be again at the same time.

This, in its turn, depends upon, and obviously derives its truth from (3):

- (3) (a) One can be at two different times at the same place.
- (b) One cannot be at two different times at the same time.

Now (3b) is obviously tautologically true, yet this tautology is created by the use of 'same' and 'different' in it, and not by anything about the nature of time itself or by the difference between time and space. One way of demonstrating this is to show that an analogous tautology can be constructed for space; viz (3c):

- (3c) One cannot be at two different places at the same place.

The kind of formulation, therefore, which seems at first sight to catch the nature of time's arrow turns out on further examination just to depend upon a concealed tautology which has nothing particularly to do with time and for which an analogy can be constructed with respect to space. If there is a difference between space and time, it cannot be brought out by the use of such expressions as 'travel' or 'return' because all such expressions contain a concealed reference to time—to travel is to change places with respect to time, to return is to be at the same place at a later time, and so on. Asymmetry between space and time will only

be demonstrated by finding a pair of expressions which do not have the kind of concealed reference to time (or place) involved in (1) and in which replacing 'place' with 'time' and 'time' with 'place' throughout results either in factual falsehood or logical incoherence (according to the type of asymmetry desired). This is the aim of the present paper.

§B *A pack of experiences*

Perhaps the asymmetry between space and time can be captured by referring not directly to objects but instead to how these objects are experienced or perceived. For perhaps the objects themselves just exist in a space-time continuum without particular directions and it is only our experience of them that gives time its (real or apparent) direction. So let us try the following pair:

(4) (a) One can experience the same place twice.

(b) One cannot experience the same time twice.

At first sight this seems no more helpful than (1) since here again 'twice' contains a concealed reference to time, meaning as it does 'at two different times'. If this reference is made overt, as happens in (5a), again it seems that the truth of (4b) depends upon a mere tautology having nothing particularly to do with time and for which an analogous tautology can be constructed for space (as in (5b) below).

(5) (a) One cannot experience the same time at two different times.

(b) One cannot experience the same place at two different places.

However, on closer inspection it can be seen that neither (5a) nor (5b) are tautologies. This is because the introduction of experience produces an ambiguity in exactly what 'time' and 'place' refer to in these expressions. They can either refer to the time and place occupied by the object of experience or to the time and place occupied by the experiencer. For example, if someone reports that he had an interesting experience in the observatory at midnight last night because he saw a solar flare, then the time and place occupied by the experiencer are quite different from the time and place occupied by the object of experience. Since, therefore, there are two different things which both 'time' and 'place' can refer to in (5a) and (5b), these two expressions are not the tautologies they appear to be, since 'time' or 'place' might refer to a different thing on each of its appearances. If, for example, the first 'time' in (5a) referred to the time of the object of experience, and the second 'time' referred to the time occupied by the experiencer, then someone could experience the same time at two different times by, for example, first seeing and then hearing a gun fire. Similarly, he could experience the same place (of object of experience) at two different places (of experiencer)

by, for example, seeing the sun from first England and then Australia. It is also possible at one time or place of experience to experience objects existing at two different times or places. So rather than (5a) and (5b) expressing tautologies, it seems that they are both quite simply false once the ambiguity in 'time' and 'place' in them is realised.

It might be thought to be more hopeful to try and expand (4a) instead of (4b) and so arrive at the following pair of candidates for displaying asymmetry between space and time:

(6) (a) One can experience the same place at two different times.

(b) One cannot experience the same time at two different places.

However, although (6a) is clearly true, the ambiguity in 'time' and 'place' between time and place of the experiencer and time and place of the object of experience means that (6b) is not true. This can again be demonstrated by example. Someone need only move between seeing and hearing a gun fire in order to experience the same time (of object of experience) at two different places (of experiencer).

If 'place' and 'time' are understood in a different sense on each occasion of their appearance in (5) and (6), then no asymmetry between space and time emerges. If they are understood in the same sense then (5a) and (5b) reduce to obvious tautologies again revealing no asymmetry between time and place. So if any asymmetry is going to be found with the aid of these formulations it must be by taking (6a) and (6b) and specifying that 'place' and 'time' in each formulation refers either to the object of experience throughout or to the experiencer throughout. However, in this case the question of whether there is an asymmetry or not becomes the question of whether something (an object of experience or an experiencer) can be at two times at the same place but not be at two places at the same time. This question is a question about objects in general and can be asked without any particular reference to experience at all; it will therefore be considered in the next section. Deal and shuffle this pack of experiences how we may, we can produce no asymmetry between space and time. If it is to be done, it will be done by the nature of objects alone.

### §C *Many body problems*

If (6a) and (6b) are stripped of reference to an experiencer or objects of experience and are rewritten so that they express a point just about objects they become:

(7) (a) A body can be at two different times at one place.

(b) A body cannot be at two different places at one time.

It might be thought that this pair of expressions exhibits the searched

for asymmetry and shows, for example, why I can return to Cambridge but not return to 1961. However, on further examination, this is not so obvious since the notion of place involved (which produces the crucial impossibility in (7b)) is multiply ambiguous and can be understood in at least three different ways.

For example, 'place' could be taken to mean 'the whole place a body occupies'. (It is understood in this way in Professor Swinburne's book, *Space and Time*, where an expression similar to (7b) is put forward as expressing a necessary truth.) However, if it is understood in this way, then (7b) again expresses a tautology showing nothing particular about time and for which an analogous tautology can be constructed referring only to space. For, on this way of understanding (7b) it becomes (8a). The spatial analogy is given in (8b):

- (8) (a) A body cannot occupy both the whole place it does and also another place at one time.
- (b) A body cannot occupy both the whole time it does and also another time at one place.

Both of these are tautologies since they both express the idea that nothing can be both everything it is in a certain respect and also something further in that respect. This way of understanding 'place', therefore, does not produce an asymmetry.

Alternatively a place could be thought to be a point, so that, on the assumption that places can be ordered in dimensions and assigned numerical values, different places will be referred to by different co-ordinates with respect to the same origin. If 'place' is understood in this way, then (7) becomes:

- (9) (a) A body can be at two different times at one co-ordinate.
- (b) A body cannot be at two different co-ordinates at one time.

Here (9a) is true, but (9b) is obviously false since any body having any size whatsoever will occupy more than one co-ordinate at any one time. So again this pair of expressions does not display an asymmetry.

There remains what is probably the most natural interpretation of 'place', and so of (7), although it is the hardest to express succinctly. This is where 'place' refers to any region completely occupied by a body at any one time for which it is the case (at that time) that a spatial path completely occupied by the body in question exists from any point in that region to any other point in that region. That is, if it is impossible to join together one point occupied by the object with another point occupied by the object without leaving the object, then the object occupies more than one place. If (7b) is understood in this sense, then we get (10b). By substituting spatial notions for temporal notions



throughout, and by substituting 'can' for 'cannot' we get (10a). If both (10a) and (10b) are true then an asymmetry will have been displayed.

(10) (a) A body can be at two different times at one spatial point between which there is not a complete temporal path occupied by that body at that spatial point.

(b) A body cannot be at two different spatial points at one time between which there is not a complete spatial path occupied by that body at that time.

At first sight this seems to display the desired asymmetry. (10a) is clearly true, as is shown by my initial example of returning to Cambridge. Here I am twice at one spatial point without its being the case that I spent the intervening times at that point. Yet if I am in two places at once (for example, both inside and outside the room), then it seems that I must also occupy the intervening spatial points (for example, be halfway through the door). So it seems that (10b) is also true.

On closer examination, however, it can be seen that there is at least one class of cases which shows that (10b) is false. This is all those things which we wish to consider as single bodies even though their microscopic parts are spatially discrete from one another. On certain physical theories, all bodies are in this position once we consider their atomic and subatomic constituents. At a more macroscopic level, such things as newspaper pictures are considered as single objects even though the parts composing them are spatially discrete. So if (10b) is taken to apply with unrestricted generality, it is obviously false. It would be possible to eliminate this class of counter-examples by specifying that (10) is only meant to apply to phenomenal objects, so that only perceptible distinctness of paths and only perceptibly continuous spatial or temporal paths need be considered. However, although it would be possible to head off this objection in this way, it seems better to reformulate (10) to show that discreteness of occupation of one spatial point or time is not the central issue. For as long as discreteness is being considered, there seems to be no significant difference between space and time. We allow single objects to have discrete microscopic spatial parts; but we also allow single objects to have discrete microscopic temporal parts (for example, a single image on a cinema screen). On the other hand, we do not allow a single object to have discrete phenomenally observable spatial parts; but neither do we allow an object to have discrete phenomenally observable temporal parts. Exceptions are often made to this in the case of people but these exceptions apply equally to both space and time. It has been held to be conceptually possible that a single person should disappear and reappear later; but it has also been held to be conceptually possible that a single person could simultaneously occupy

two different spatial positions (e.g. be both here and in Australia at once). What does seem to establish an asymmetry, and what gives the impression that my returning to Cambridge is something for which no temporal analogy can be given, is not the fact that some point, spatial or temporal, is occupied in a discrete manner but that in the one case continuity can be provided by the occupation of other points in the intervening times and in the other case no such continuity can be provided.

In the case of the return to Cambridge, that is, the key point is the continuity of my existence between the two times I am at Cambridge. This is guaranteed by my occupation of other places during the intervening time. It seems impossible, however, that I could exist both here and in Australia in a way which would allow an analogous kind of continuity; it does not seem that this would be provided by my occupation of the intervening places at other times. We can use one aspect of this feature of continuity in order to provide the following pair of formulations which should therefore exhibit the required asymmetry:

- (11) (a) A body can be at two different times at the same spatial point when it is the case that, for any intervening time, it occupies a spatial point at that time and not all these spatial points are the same.
- (b) A body cannot be at two different spatial points at the same time when it is the case that, for any intervening spatial point, it occupies a time at that point and not all these times are the same.

The claim involved in these formulations is probably clearer if it is thought of as a claim about which kinds of graphs could represent possible occupations of space and time by a single body. If time is thought of as being plotted up and down the page and space across the page, then the claim is that a graph of the shape of a left or right hand bracket,  $\langle$ ,  $\rangle$ , shows possible occupation but that a graph of the shape of a French circumflex or of a V does not,  $\wedge$ ,  $\vee$ .

While this may reflect a factual asymmetry, and so explain why I, living in the world that I do, have to try and cure my nostalgia by a second occupation of a place rather than a second occupation of a time, I do not see why it should be taken to show a conceptual asymmetry. The symmetry of the graphs shows that the one type of occupation of space and time is as easy to describe as the other. It just seems to be a fact (at least on the macroscopic level) that we do not experience objects which exhibit the kind of behaviour which would be represented by the graphs,  $\wedge$ ,  $\vee$ . On the other hand, if our experience were to change, or if certain theories about the behaviour of subatomic particles were to

gain ground, so that situations such as are represented by these graphs became common, then there seems no reason why we should not change our concepts and hold these to be occupations of space and time by a single object. If we did, then (11b) would be false and there would no longer be any asymmetry between space and time.

*University of Bristol.*

## CONFIRMABILITY AND FACTUAL MEANINGFULNESS

*By R. G. SWINBURNE.*

I WISH in this article to examine a principle which I shall call the confirmationist principle. This is that a statement  $q$  is factually meaningful if and only if either  $q$  is itself an observation-statement or there are observation-statements which, if true, would confirm or disconfirm  $q$ . By 'confirm' I mean 'raise the probability of'; by 'disconfirm' I mean 'lower the probability of'; by an 'observation-statement' I mean a proposition which reports a state of affairs which it is logically possible to observe. I understand by a 'statement' whatever is expressed by an indicative sentence. I understand by a statement's being 'factually meaningful' that it describes a logically possible state of affairs, which, it makes sense to suppose, might or might not hold of the world.

The confirmationist principle expresses (in slightly different terminology) the very weak form which verificationism reached in *Language, Truth and Logic*. There is no claim here about the meaning of propositions *being* their method of verification, nor any claim about propositions' needing to be *conclusively* verifiable in order to be factually meaningful; nor any claim about factual meaning being the *only* kind of meaning. These stronger claims were quickly abandoned for well known reasons, but the weaker claim of the confirmationist principle remains with us. Ayer and others have devoted much subsequent attention to the issue of when one statement confirms another (or, as Ayer would say, 'verifies' another 'directly' or 'indirectly'). Is it sufficient for  $p$  to confirm  $q$  that  $q$  entail  $p$  and not entail  $\sim p$ ? Attempts to lay down criteria of confirmation on these lines have run into considerable difficulties. (See, for example, the discussion in James W. Cornman, 'Indirectly Verifiable: Everything or Nothing', *Philosophical Studies*, 1967, 18, 49-55). But even if it is not possible to give a general account of when one statement confirms another, that has no tendency to cast doubt on the applicability of the confirmationist principle itself.

So long as we can in general recognize intuitively when one statement confirms another (as of course we can) we can apply the principle.

The discussion of attempts to fill out the confirmationist principle by giving a general account of when one statement confirms another has directed attention away from the principle itself. In this paper I re-examine the principle. I shall maintain that the confirmationist principle is false because there are factually meaningful statements which are neither observation-statements nor confirmable or disconfirmable by observation statements; and, further, that even if the principle were true, it would not be of great value in sorting out factually meaningful statements from others.

Let us begin with the second point. In order to use the confirmationist principle to sort out sense from nonsense we will have to have an understanding of which statements are observation-statements. By 'an observation statement', it will be recalled, is meant a statement which reports a state of affairs which it is logically possible to observe. Because of uncertainty about whether a statement was "factually meaningful", that is, described a logically possible state of affairs, the confirmationist principle was proposed that a statement described a logically possible state of affairs if and only if either it described a logically possible state of affairs which was observable or it would be confirmed or disconfirmed by some statement which described a logically possible state of affairs which was observable. Hence the principle will be useful only if men are agreed better about which statements report observable logically possible states of affairs than about which statements report logically possible states of affairs in general. Are they? I doubt it. Consider all the things that some men have claimed to be observable: The end of the world, one's own death, the devil, Heaven, the fourth dimension, Poseidon, men turned into stones, etc., etc. Some men have held these things to be observable "in principle" and others have denied it. Is there any simple way to settle the issue? One way which has been suggested is to suppose that the observable must be describable by some simple sensory vocabulary; to claim that we can *really* observe objects which are square or round, red or blue, move, utter noises, etc., but that we cannot observe men who are quarrelsome, a lump of gold, or the planet Venus. When we claim to have observed the latter, we ought rather to claim to have inferred these things from things of the former kind which we can truly be said to have observed. But this kind of move constitutes a highly arbitrary restriction on the normal use of 'observe'. We now recognize that there is no simple and obvious limit on the observable—we can observe bacteria (under a microscope), the moons of Jupiter (in a telescope), Mr. Nixon (on the television). So if men claim to have observed the objects and events cited earlier how can we show them wrong? One or two of the purported observations *might* be eliminable

in virtue of a logical property of the word 'observe' such as that it is not logically possible to observe the future. From this it would follow, for example, that an inhabitant of the world cannot observe its end. For many  $p$ , however, the proof, if it can be given, that  $p$  is not something which it is logically possible to observe, will, I suspect, consist of a proof (by means other than the confirmationist principle) that  $p$  is not logically possible *simpliciter*. So although men may be agreed *by and large* about which statements are observation statements, I see no reason to suppose that the degree of consensus is vastly greater here than over which statements are factually meaningful. And if that is so, the confirmationist principle is not going to be of great help in clearing up the latter.

Even if men were to agree for any given statement about whether or not it is an observation-statement there is a second difficulty which arises with any attempt to show that the statement expressed by some sentence is *not* a factually meaningful proposition. To show this you have to show that it is not confirmable or disconfirmable by *any* observation statement. Yet we hardly have before us a catalogue of types of observation-statement which we can run over quickly to see whether they have any confirmation relations to some given statement. So we may easily make a mistake in concluding that a statement  $p$  is not confirmable or disconfirmable by any observation-statement through not having thought of a certain observation-statement  $q$ , which does in fact confirm (or disconfirm)  $p$ . Many a scientific theory when first put forward must have met the reaction—how on earth could you show this to be true? But later ways were found. Consider the early particle theory of light—the claim that light consists of tiny particles. An immediate reaction to this is to say 'How on earth could this be confirmed?' We cannot observe light, for light is that by means of which we do our observing. But later we learn about the photo-electric and Compton effects and come to see that they confirm the particle theory.

For these reasons, but primarily the first, I conclude that the confirmationist principle may not be of much use in separating the "factually meaningful" from what is not "factually meaningful". Nevertheless the principle may be true. Is it? I do not know of any very good argument in its favour. Sometimes proponents argue from examples. Certainly many statements which we judge to be factually meaningful appear on reflection either to be observation-statements or to be confirmable or disconfirmable by such. But that does not prove that there are no exceptions. A vague argument is sometimes put forward to the effect that we could not understand a factual claim unless we knew what it would be like to observe it to hold or knew which observations would count for or against it; from which it follows that a statement could not be factually meaningful unless it was an observation-statement or there

could be observational evidence which would count for or against it. But surely we understand a factual claim if we understand the terms which occur in the sentence which expresses it, and if they are combined in a pattern of which we understand the significance. It may be that in order to understand the terms we have to have observed instances of their correct application or instances of the correct application of terms definitionally related to the former—or at least to have observed events which are evidence for or against the occurrence of such instances. And it may be that in order to understand the significance of the pattern in which the terms are combined (e.g. a subject-predicate sentence) we have to have observed instances of the correct use of such a sentence-pattern. But none of this shows that in order to understand a particular statement we have to know what it would be like to observe *it* to hold or know which observations would be evidence for or against *it*. And such an example as that of the particle theory of light cited above suggests that we could understand a proposition *without knowing* what it would be like to observe it to hold or which observations would count for or against it.

I conclude that arguments in favour of the confirmationist principle are less than successful. I wish now to argue that the principle is false because conclusive counter-examples can be provided of propositions which are factually meaningful although they are neither observation-statements nor confirmable or disconfirmable by such. The propositions which I shall cite are in various ways different from each other, but have a common form. They all assert the existence of a discrepancy between the best available evidence which men will ever obtain about whether some state of affairs *S* holds and whether or not *S* holds. Such a proposition will be factually meaningful if the claim that *S* holds is not the same as the claim that best evidence men will ever have supports the claim that *S* holds to such and such an extent. For example:

$p_1$ : Among possible claims about the prehuman past which the best evidence ever to be obtained by man makes highly improbable some are nevertheless true.

By 'the prehuman past' I mean 'what happened before the existence of men'. (I understand the term 'man' here and hereafter to cover any being capable of assessing evidence.) At any time man has a certain amount of evidence about what happened before the existence of men, and the best evidence which he can get is going to make some claims about the past very improbable. What we know now makes it very improbable that an empty rocket took off for the moon from Cape Kennedy in the year 500 million B.C. Yet clearly it is a different thing to say that some past event did not occur from saying that some later evidence makes its occurrence very improbable. So  $p_1$  is factually

meaningful—at any rate given the plausible supposition that there is a prehuman past; if there is not it will fail of reference. Some of our best evidenced beliefs about the past may be false. Yet nothing men could ever observe could count for or against  $p_1$ . If we have the best evidence which men will ever have about some proposition  $q$  making a claim about the prehuman past and this evidence makes  $q$  very improbable, then we cannot add to it any additional evidence about whether  $q$  is true. If we *know* that a possible claim about the prehuman past is correctly classified as ‘a claim which the best evidence ever to be obtained by man makes highly improbable’, then we cannot, *while* knowing that that description applies, obtain any further evidence about whether  $q$  is true. Of course we may not know that some claim  $q$  is so classifiable; we may suspect that we may get more evidence which will put us in a better position than we are today to assess the truth-value of  $q$ . Yet in so far as our present evidence renders it likely that  $q$  is ‘a claim which the best evidence ever to be obtained by man makes highly improbable’, it will *ipso facto* render it probable that  $q$  is false. Nor can there be any other evidence relevant to the truth or falsity of  $q$ . Whatever confirms a proposition confirms that the best evidence ever to be obtained will confirm that proposition. (Evidence that there was an earthquake in Chile in 500 m. B.C. is evidence that the best evidence ever to be obtained will show that there was an earthquake in Chile in 500 m. B.C.). So there cannot be any evidence which counts for any claim  $q$  being a claim which the best evidence ever to be obtained makes highly improbable and at the same time counts for  $q$  being true. From this it follows that no evidence can count for  $p_1$ . For if the probability of any claim  $q$  about the prehuman past both having the former property and being true is the same on any evidence, then necessarily the probability that there is a claim which has the former property and is true is the same on any evidence. Nor can any evidence of observation count against the truth of  $p_1$  for the same reason that no evidence of observation can count against ‘Some bachelors are married’. We know in advance of observation that any evidence which turns up showing a man to be a bachelor will be evidence against his being married. Whatever we subsequently observe cannot add to our grounds for believing ‘Some bachelors are married’ to be false. Similarly with  $p_1$ . We know before making any observations that whatever evidence confirms any claim  $q$  about the prehuman past to be ‘a claim which the best evidence ever to be obtained by man makes highly improbable’ disconfirms  $q$ . So whatever observations we then make, they are not going to add to our grounds for believing  $p_1$  false. Yet, it makes sense to suppose that  $p_1$ , unlike ‘All bachelors are unmarried’, has either truth-value. I conclude that nothing which men could ever observe could count for or against  $p_1$ . So  $p_1$  cannot be an observation-statement (since nothing observable could

## IS THERE AN OZMA-PROBLEM FOR TIME?<sup>1</sup>

By NORMAN SWARTZ

THE problem derives from Kant [4, §13], but the name by which we currently refer to it is due to Frank D. Drake [2, pp. 158–159]. William James stated the Ozma problem in this fashion:

If we take a cube and label one side *top*, another [presumably the side parallel to it] *bottom*, a third *front* and a fourth [again, presumably, the side parallel to the latter] *back*, then there remains no form of words by which we can describe to another person which of the remaining sides is *right* and which is *left*. [3, Vol. II, p. 150, parenthetical remarks added.]

Martin Gardner puts the problem explicitly as one of communication:

... Is there any way to communicate the meaning of 'left' by a language transmitted in the form of pulsating signals? By the terms of the problem we may say anything we please to our listeners, ask them to perform any experiment whatever, with one proviso: *There is to be no asymmetric object or structure that we and they can observe in common*. [2, p. 160.]

The Ozma problem has never been raised in regard to any but spatial relations. Is this just lack of imagination? Richard Taylor [5] has argued, I think persuasively, for a very much more thorough analogy between spatial and temporal relations than had previously been allowed. How much further can we press the analogy? Specifically, can we construct an Ozma problem for time?

The task, it would seem, divides into two stages. First we must formulate the proper analogy, and second we must investigate whether the problem is soluble.

Taylor proposes a one to one mapping of temporal terms into spatial ones. 'Now' and 'here' are correlated; so too are 'a time' and 'a place'; similarly 'future to' and 'north of'. (Purely for convenience we shall prefer to correlate 'later than', an obvious near-synonym of 'future to', with 'left of' rather than, as Taylor does, with 'north of'. This change is of no especial significance. It merely points up the fact that while there are three spatial directions there is only one temporal one; consequently the temporal direction is mappable onto any one of the three spatial directions indiscriminately.)

For current purposes, Taylor's list is insufficient. We shall require a few more terms in our dictionary. If we wish to translate Gardner's statement from spatial-talk into temporal, we shall have to have dictionary equivalents for both 'pulsating signals' and 'asymmetric object'.

<sup>1</sup> The author acknowledges his indebtedness to J. F. Bennett's 'The Difference Between Right and Left' [1]. That this paper can be as brief as it is is due to the thoroughness of Bennett's work.



I take 'pulsating signals' to mean something like 'a series of markers (for example audible beeps and their absences) arranged in a temporal order'. Implicit in Gardner's story is that the signals or messages should be received during a time, i.e., consecutively, but at one place. A Taylor-analogue could be something of this sort: 'a series of markers arranged in a spatial order'. To insure the completeness of the analogy we should have to add the rider that this message should be received with a spatial extent but at one time. A notched iron bar, for example, could satisfy the description. The message is extended over a space (i.e. the length of the bar) and is all of it simultaneously present.

The temporal analogue of a spatially asymmetric object is easy to name but perhaps slightly more difficult to explain. A temporally asymmetric object is one for which there is no moment such that the history of the object up to that moment is the "reflection" of the future of that object subsequent to that moment. Noticeably the explanation seems to be infected with a slight remnant of space-like talk, to wit, 'reflection'. The term is, however, a mere convenience. We have far more occasion to speak of spatial reflections than of temporal ones, but the suitability of the term for both contexts should be all too obvious. Most objects are temporally asymmetric, but not quite all. With a little ingenuity we can actually make a temporally symmetric object. For example, a pure pitch of constant volume, physics tells us, would sound exactly the same if recorded and played backwards. There is in this example a temporal axis of symmetry. There is some moment, the mid-time of its duration, about which the sound is temporally symmetric, i.e. reflected.

It is important to note for our discussion below that physical objects, too, are very often temporally symmetric within selected time intervals. For example, the notched iron bar which we have alluded to, may for years undergo no internal change whatever. During that segment of its total history it is temporally symmetric. If nothing is happening to it over a course of, let us say, two centuries, then the description we give of what happens to it from the first year to the end of the one hundredth year is precisely the same as any description we give of it from the end of the two hundredth year (backwards) to the beginning of the one hundred-and-first year. Putting the matter perhaps a bit cryptically: we cannot tell by examining the bar during that two-hundred-year interval whether it is growing older or younger. In this respect it is the temporal analogue of a spatial object which has no preferred direction, e.g. William James's cube.

For fairly obvious reasons in the statement of the temporal Ozma problem only temporally symmetric objects are to be mutually accessible to both of the two worlds which are in communication. Allowing a bit of judicious editing, the problem emerges thus:

Is there any way to communicate the meaning of 'earlier than' by a message transmitted all at once by a series of spatially arranged markers? By the terms of the problem we may send any message we please, with one proviso: There is to be no temporally asymmetric object or structure to which the sender and receiver have mutual access.

Does this problem really make sense? Can there be a problem about communicating the meaning of 'earlier than' at all analogous to the problem of communicating the meaning of 'left of'? We understand how one person communicating with another by telephone should be frustrated in trying to tell the other person, who did not already know, which was his left hand and which his right. Could there really be a problem in trying to instruct the other what the difference between earlier and later is? The very fact that they are communicating at all, that one is saying *now* this and *then* that, would, along with the use of some token-reflexive expressions, seem to solve the problem. Even if by the terms of the problem, the message must be received in its entirety all at once so that conversation between the two parties is implicitly excluded, the *reading* of the message, unlike the mere receiving of the message, is emphatically *not* instantaneous but must take some amount of time. We can imagine the sender forwarding the following sort of message:

*Dear reader: I should like to explain to you how I am using the terms 'earlier' and 'later'. The reading of this message, we both know, takes time. We further know that events are ordered in time. You can understand what I mean by 'earlier' and 'later' by reflecting on the point that as I use the term you will have read the word, 'Dear', earlier than you will have read the word 'reader'. Later is the converse relation to earlier.*

This ready solution to the temporal Ozma problem is, however, mere illusion. The argument harbours an important implicit assumption which, when exposed, undermines the conclusion. To reveal this assumption I propose to pursue the problem in a slightly altered, but equivalent, way.

Let us imagine the original spatial version of the Ozma problem as concerning not the instructing of a remote listener as to which is right and which left, but rather the attempt to discover which items in the two worlds, linked in communication as we have described, are incongruous counterparts or enantiomorphs. The two versions of the spatial Ozma problem are equivalent in this respect: a solution to either one would provide a means to solve the other. If we knew that the person with whom we were talking on the telephone had a body spatially congruent to our own, i.e. was not enantiomorphic relative to us, we could simply tell him that his right hand was the one which was on the same side of his body as his appendix. The solution to the first problem

provides a means of solving the second, and (it should be obvious) conversely.

In the light of this alternative description of the spatial version of the Ozma problem, we can construct a second formulation of the temporal Ozma problem. Under this revision we can conceive of the problem, equivalently, as concerning the attempt to inform the receiver whether the region in which the message originated shares a time direction congruent with that of the receiver or whether the two regions are temporal enantiomorphs of one another.

Stating the problem in this way, we can readily see what is wrong with the suggested solution to the temporal Ozma problem: it helps itself gratuitously to one of the two possibilities which are to be decided between. For consider: if the region in which the message is received is suffering a local time reversal, then the test forwarded by our misguided sender will have precisely the opposite effect to the one intended. If we were to observe that strange world directly it would appear to us to be running backwards. Persons would grow younger instead of older, golf balls would fly out of holes and stop abruptly at the head of a putter, which is then cocked upward, etc. Similarly for the reading of the message. Relative to *our* time, the receiver would "start" reading at what we take to be the end of the message and would finish reading some time later with the word 'Dear'. 'Dear' would be read later than 'reader', not earlier.

The analogy with the spatial Ozma problem thus emerges rather more live than the first solution would have led us to think it might. But if this solution does not work, is there any that does? Is there any message conceivable, subject to the constraints given, which would allow the passing of decisive information concerning the relative directions of time in the regions of the sender and of the receiver? The analogous spatial Ozma problem, the problem of finding out whether our correspondent is like or oppositely handed, is generally acknowledged to be *insoluble* [1]. Is the temporal Ozma problem similarly insoluble?

An argument can be given that it is similarly insoluble. (In it we imagine ourselves in the position of receivers of the message.)

Suppose we were able to observe directly two worlds which are temporal enantiomorphs of one another. (If we were to film the histories of both of these worlds and then run *one* film backwards, the scenes projected would be qualitatively identical.) In one of these worlds, the one in which time runs the same way as ours, we see a man pick up an iron bar, clamp it in a vice, and studiously proceed to file a series of notches in it. For a while filings fall to the floor. At last he is finished, he unclamps the bar, and places it away on a shelf where it sits for a very long time.

As we look in on the second world, we see a similarly notched bar sitting on a shelf. At first all appears normal. After a while, however, strange things begin to happen. A man walks backward into the room, takes the bar from the shelf and clamps it in a vice. A file rises abruptly from the workbench to his waiting hand. He places the edge of the file into one of the notches in the bar. Suddenly some filings leap from the floor to meet the file which is abruptly drawn across the bar, welding these flying particles into the notch in the bar. And so the story goes on.

But what about the bars during these longish periods when they sit quiescently on their shelves? If we look at one of the bars during just that time we cannot tell in which of the two worlds it resides. We must wait to see what happens *in that world*. The Ozma problem asks us to imagine such a bar out of its world, to imagine it wrenched from its surroundings with their tell-tale clues and delivered naked to us. Can we tell by examining the bar which world it originated in? The conclusion we are driven to is that there is *no* way. Anything that the writer in the first world could have written could, with equal likelihood, appear verbatim in a message from the second world. We would be totally at a loss to assign such a message to one world or to its temporally reversed counterpart. Thus the analogy between the spatial and temporal versions of the Ozma problem seems complete.

Surprisingly, this conclusion too, just as its erstwhile opposite, follows from a defective argument and must be rejected. Let us see why.

This latter argument, an argument for the insolubility of the problem, assumes that we can transport the iron bars in question out of each of the two worlds described. There is, of course, no problem for the world in which the direction of time is the same as that in our world. We can imagine the machinist in the first case returning eventually, taking the bar down from the shelf, ensconcing it in a rocket, and launching the rocket heavenward. Years later the rocket in its flight through the void is intercepted quite accidentally by some Earthlings who, knowing nothing of its place of origin, take its cargo home to study.

But what story shall we tell for the second case? What sort of *causal* chain of events can deliver a bar to us from a world where time is oppositely directed? There would seem to be a profound difficulty in there being communication, or causal interaction, between two worlds so related. For the second world the rocket-ship account simply will not work. For the man in our second story to take the bar, place it in a rocket, and launch it off into space, would *for him in that world* be a case of his operating contrary to the laws of thermodynamics. We would expect in this second world that rocket ships would land while sucking in flames and smoke, but not that they should lift off spewing out flames and smoke.

How then *are* we to get possession of the bar?

It would appear that the only way for us to receive the described sort of message from a world whose time direction was opposite to ours would require that there be a violation of some causal laws in one or both of the two worlds.

So far-reaching is this point that it even requires that we retract the *Gedankenexperiment* which originally set the stage on which the problem was to be played out. We have uncritically imagined someone looking in on both of two worlds having opposite time directions. But even this amount of causal interaction is in violation of causal laws. Part of the story we tell of the process of seeing involves the emission of photons from objects and the *subsequent* impinging of these photons on our retinas. But this process is obviously directed in time. In a world where time ran oppositely to ours, we could not see objects at all: objects would be photon-sinks, not photon-emitters.

In sum, the analogy ultimately flounders. The Ozma problem for spatial relations is genuine in this world as matters now stand. No causal laws need be violated for two spatially incongruous worlds to be in temporal communication with one another. The analogous situation does not hold for temporally incongruous worlds. For two temporally incongruous worlds to be in communication by means of a spatially extended message does require the violation of causal laws in one or both of the two worlds concerned.

Taylor-type analogies between spatial and temporal relations do have their limits. The Ozma problem is one feature of spatial relations which is without counterpart among temporal relations.

*Simon Fraser University*

## REFERENCES

- [1] Bennett, Jonathan F., 'The Difference Between Right and Left', *American Philosophical Quarterly*, Volume 7 (1970), pp. 175-191.
- [2] Gardner, Martin, *The Ambidextrous Universe*, New York, 1964, especially chapters 18-25.
- [3] James, William, *The Principles of Psychology* (original edition, 1890). Present references are to the Dover Edition, 1950.
- [4] Kant, Immanuel, *Prolegomena to any Future Metaphysic that will be able to present itself as a Science*, translated by P. G. Lucas, Manchester, 1953.
- [5] Taylor, Richard, 'Spatial and Temporal Analogies and the Concept of Identity', *Journal of Philosophy*, Volume LII (1955), reprinted in *Problems of Space and Time*, ed. J. J. C. Smart, New York, 1968.

## ARE COGNITIVE PROCESSES AND STRUCTURE A MYTH?

By MICHAEL MARTIN

NORMAN Malcolm has argued that cognitive processes and structure are a myth, that psychologists explain nothing by their theories of cognitive processes and structure since there are no such processes or structures.<sup>1</sup> In this paper we will show that Malcolm's arguments for this position are unsound.

### *The Alleged Myth of Cognitive Processes*

Malcolm argues:

There is a real difficulty . . . in our understanding what a psychologist is talking about when he says, for example, that he wants to explain "the processes of pattern recognition" or to construct a model for it. In recognizing patterns, shape, colors and people there is usually or often no *process* of recognition. So *what* is the model a model of?

For example, there is often no process involved in recognizing someone's face. You see a friend's face in a crowd. 'You smile at him and say "Hi John"'. You do not think "Where have I seen that face before?"' The same thing is true of remembering. Sometimes one does go through a process of *trying* to remember, according to Malcolm. For example you may "review in your mind" the places you went on the way home in order to recall where you left your briefcase. But in most cases there is no review. We just remember where, for example, we have left the pliers or where our clean shirts are.

It is important to note that Malcolm admits that sometimes there are cognitive processes, e.g. in the case of trying to remember. He maintains, however, that these cases are rare. Now, in the cases of cognitive processes Malcolm acknowledges, he seems to be talking about cognitive processes that are easily accessible to introspection. In his example you review in your mind where you went on the way home. The processes that psychologists are talking about, however, need not be processes like that.<sup>2</sup> Psychologists might well admit that processes of the kind Malcolm is talking about occur only sometimes. This admission would be perfectly compatible with the thesis that processes that are unconscious or subconscious always or at least often occur. Moreover, whether the

<sup>1</sup> Norman Malcolm 'The Myth of Cognitive Processes and Structures', *Cognitive Development and Epistemology*, ed. Theodore Mischel (New York: Academic Press, 1971), pp. 385-392. Malcolm's paper is in part a response to P.C. Dodwell's 'Is a Theory of Conceptual Development Necessary?', which also appears in this volume (pp. 365-388). Dodwell's response to Malcolm is in a postscript. Dodwell's response, although correct, does not attempt to evaluate Malcolm's specific arguments.

<sup>2</sup> For some typical work in cognitive psychology see J. Bruner, J. Goodnow and G. Austin, *A Study of Thinking* (New York: John Wiley, 1956) and J. Bruner *et al.*, *Contemporary Approaches to Cognition* (Cambridge: Harvard U. Press, 1957).

cases Malcolm acknowledges are as rare as Malcolm seems to suppose is not clear. Certainly Malcolm gives no evidence to support his thesis on this point. Even if these cases are as rare as Malcolm says, some do by his own admission exist. Such processes can hardly be myths in the cases where they do exist, and psychologists surely in these cases have a subject for their theories and models. At most, then, Malcolm has shown that psychologists' models of cognitive processes do not have as wide an application as they are thought to have, not that they have no application and hence are models of nothing. However, as I have suggested, psychologists have in general not been talking about conscious or introspectible cognitive processes.

Malcolm does not just lean on introspective evidence to establish his thesis. He argues that appeal to inner processes is a *philosophical* mistake. For the supposition that there is an inner process of recognition is based on an incorrect view of mind. The correct view does not appeal to inner processes; yet it is not strictly speaking behaviouristic. For example, recognition of John's face is not an inner process. To be sure, there may be all sorts of images, feelings, thoughts that are associated with recognizing a face. However, even if there are not, recognition is still possible. The criterion of whether someone has recognized a face is not in terms of the person's images, feelings or thoughts. This does not mean that recognition consists just in the behaviour of saying 'Hi John'.

Imagine an eccentric who smiles at and says 'Hi John' to every tenth person he passes; and who has never seen this John before. Given those facts, his smile and utterance on this occasion would not be an expression of recognition. On the other hand, it is easy to imagine a situation in which such a smile and greeting would be an expression of recognition. Thus, it is the facts, the circumstances surrounding that behavior, that give it the property of expressing recognition. This property is not due to something that goes on inside.<sup>1</sup>

Thus, according to Malcolm, recognizing John's face is *appropriate behaviour in certain circumstances*.<sup>2</sup>

Several critical points can be raised about Malcolm's analysis. First, one might want to make a distinction between the achievement and process meaning of terms.<sup>3</sup> The term 'learning', for example, sometimes refers to certain achievements of people and sometimes to the process leading up to such achievements. It might be argued that 'pattern recognition' and 'remembering' also have an achievement and a process sense. Malcolm might then be interpreted as claiming that pattern recognition and remembering in the *achievement sense* refer only to certain behaviour in certain contexts, and he might well be correct. But this would show nothing about what remembering and pattern recognition

<sup>1</sup> Malcolm, *op. cit.*, p. 387.

<sup>2</sup> Malcolm is, of course, indebted to Wittgenstein for this view.

<sup>3</sup> I am indebted to David Marion for this point.

are in the *process sense*. In this sense inner processes might indeed be relevant for the analysis.

But let us suppose that there is no process sense of 'pattern recognition' or 'remembering', or, perhaps, that the analysis of the process sense of 'pattern recognition' or 'remembering' does not refer to anything inside the person who recognizes the pattern or who remembers. Still one might want to know *why* Jones' face was recognized and the explanation of this may refer to inner processes. Thus there is an important distinction to be drawn between two theses:

- (1) A philosophical analysis of the terms 'pattern recognition' and 'remembering' in either the achievement or process sense involves inner process terms.
- (2) A scientific explanation of the phenomena of pattern recognition and remembering involves reference to inner processes.

Malcolm seems to blur the distinction between (1) and (2). Once these theses are distinguished it is clear that Malcolm has at most refuted (1); his argument does not even seem to touch (2). Furthermore, it seems unlikely that psychologists are particularly interested in (1), if (1) is understood as giving a philosophical analysis of the commonsense or ordinary meaning of 'pattern recognition' or 'remembering'. Psychologists, like all scientists, redefine terms to suit their theoretical purposes. Cognitive psychologists have of course advocated (2). But, since Malcolm's argument does not touch (2), even if it is valid it seems to have little relevance to psychology.

### *The Alleged Myth of Structure*

Malcolm criticizes social scientists like Chomsky and Lenneberg who appeal to systems of rules of abstract structures that explain the learning of language and other types of learning. Again he argues that appeal to such rules or structures is mythological.

Malcolm first quotes a linguist R. A. Hall who disagrees with Chomsky. However, such recourse to authority is hardly cogent unless the authority has good reason for *his* disagreement. Unfortunately the quotation from Hall used by Malcolm indicates that Hall's reasons are very weak indeed. For example, Hall says:

As anyone can see by direct observation of ordinary people's normal every-day speech-activity, people simply do not talk according to rules, whether they be aware of the rules' existence or not. There are too many instances of normal speech which cannot be accounted for by any generative rules at all—which constitute, in fact, all kinds of "violations" of rules—but which cannot be neglected or whose existence cannot be denied in that account.<sup>1</sup>

<sup>1</sup> R. A. Hall, 'Some Recent Developments in American Linguistics', *Neuphilologische Mitteilungen*, 70, 1969.



It is as absurd to say that the existence of the rules Chomsky postulates can be refuted by 'direct observation' as it is to say that the existence of electrons can be so refuted. The postulation both of electrons and of these rules is based upon very indirect evidence. Ultimately the acceptance or rejection of these rules and also the acceptance or rejection of electrons will be in terms of their predictive capacity and their explanatory power. To suppose otherwise is to miss the entire point of their postulation.

Furthermore, to suppose that the existence of violations of grammatical rules in everyday speech counts against the theory is to misunderstand the theory completely. Chomsky wants to explain linguistic competence, not actual linguistic performance. Chomsky's theory purports to explain 'the behaviour of an ideal speaker, having perfect command of syntax and morphology, and free from . . . memory lapses, slips of tongue and other irregularities'.<sup>1</sup> As is well known, idealization is an acceptable and crucial part of scientific theorizing. To argue against Chomsky in the way Hall does is thus to miss the point. One might as well argue against the laws of falling bodies by pointing out that a feather and an iron ball released from a second storey window will not hit the ground at the same time. These laws and Chomsky's theory apply only under certain idealized conditions.

Clearly then Malcolm's appeal to Hall as a linguistic authority lends no support to his thesis that cognitive structure is mythological. However, Malcolm has another argument, one which purports to show that the assumption of cognitive structure leads to 'an infinite regress, or else it leaves one with the same sort of "mystery" that led to postulating of a system or structure in the first place'.<sup>2</sup> His argument seems to be this: Cognitive psychologists postulate a system of rules  $R_1$  to explain why people know how to perform some activity  $A_1$  (speak a language, recognize a pattern). However, this does not explain how people know how to  $A_2$  (use this system of rules  $R_1$ ). Hence either (a) another set of rules  $R_2$  needs to be postulated to explain why people know how to use  $A_2$  and so on *ad infinitum*, or (b) people "just know how" to  $A_2$ . But if (b) one might as well say that they just know how to  $A_1$ .

The first thing to notice about Malcolm's argument is that it is not strictly speaking an infinite regress argument. We are not forced by the argument to go back and back forever. There is an important distinction to be drawn between two schemata:

- (1) In order to explain why X knows how to  $A_n$  one must postulate a set of rules  $R_n$  that X knows how to use, where this using is symbolized by ' $A_{n+1}$ '.

<sup>1</sup> Max Black, 'Comments on Noam Chomsky's Problems of Explanation in Linguistics', *Explanation in the Behavioural Sciences*, ed. Robert Borger and Frank Cioffi (Cambridge University Press, 1970), p. 453.

<sup>2</sup> Malcolm, *op. cit.*, p. 391.

- (2) In order to explain why X knows how to  $A_n$  in terms of a set of rules  $R_n$  one must explain why X knows how to use  $R_n$  i.e. how to  $A_{n+1}$ .

Substituting in schema (1) does not lead to an infinite regress; what it leads to is an infinite series of explanations of the same sort (rule explanation).<sup>1</sup> However, in order to explain why X knows how to  $A_1$  it is not necessary to *explain* why X knows how to use a system of rules  $R_1$  ( $A_2$ ) and so on. On the other hand, in schema (2) this is precisely what one has to do. One is forced into the position of not being able to explain one activity  $A_1$  unless one can explain another activity  $A_2$  and so forth *ad infinitum*. Clearly cognitive psychologists need not be committed to schema (2) in their explanatory activity.

Indeed, it is clear that they are not even committed to schema (1). They need not suppose that the explanation of why people know how to use certain rules postulated by them (in order to explain why people know how to recognize patterns, etc.) is itself explained by further postulation of rules people know how to use. Nor do psychologists have to say that people just know how to use such rules. Malcolm's argument here is based upon a false dichotomy: Either (a) an explanation of why people know how to do something is in terms of rules they know how to use, or (b) an explanation of why people know how to do something is 'They just know how'. Psychologists might attempt to explain why people know how to use such rules by neurological theories.<sup>2</sup> Such an explanation is presumably neither a rule explanation nor a 'They just know how' explanation.

Furthermore, psychologists might say that at the present time they don't know why people know how to use the rules they postulate. There is nothing scientifically suspect in this move. Any explanation in science that answers some question may raise another question that may be impossible to answer at that time, e.g. the question what is the explanation of this explanation. However, this does not show that no progress has been made. So long as the explanatory hypothesis meets the standard methodological requirements of testability, non-ad hocness, predictive power and so on and withstands critical test, progress has been made. Whether the explanatory hypotheses of the cognitive psychologists meet these standards is a big question.<sup>3</sup> However Malcolm's paper does not even attempt to answer this question.

<sup>1</sup> For an extended account of infinite regress arguments where this distinction is made, see John Passmore, *Philosophical Reasoning* (London: Gerald Duckworth and Company, 1961), Chapter 2.

<sup>2</sup> To be sure Malcolm has argued against these types of explanation. See Norman Malcolm, 'The Conceivability of Mechanism', *Philosophical Review*, 1968, pp. 45-72. However, his argument seems to have little merit. See Michael Martin, 'On the Conceivability of Mechanism', *Philosophy of Science*, 1971, pp. 79-86.

<sup>3</sup> For an illuminating critique of cognitive theories, see Robert Schwartz's review of J. Bruner, R. Oliver and P. Greenfield *et al.*, *Studies in Cognitive Growth*, in *Journal of Philosophy*, 65, 1968, pp. 172-179.

We may conclude, I think, that although cognitive processes and structure may be myths, Malcolm has not shown that they are.

*Boston University*

## ON PROMISING THE UNWANTED

By W. R. CARTER<sup>1</sup>

JOHN R. Searle says that if person *S* 'sincerely and non-defectively promises' to person *H* that he will do *A* then '*H* would prefer *S*'s doing *A* to his not doing *A*'. Searle further claims that '*S* believes *H* would prefer his doing *A* to his not doing *A*' (*Speech Acts*, Cambridge, 1969, p. 58). Recently David Armstrong endorsed a similar, but somewhat more modest, condition for sincerely promising.<sup>2</sup> According to Armstrong, 'a sincere promise involves the speaker in believing that his audience would prefer the fulfilment of the promise to its non-fulfilment'. This seems to amount to accepting only the second conjunct of the condition posed by Searle above. When Armstrong refers, to a promisor's 'audience' I will take him to mean the person or persons to whom the promise is made, and not anyone who happens to overhear the promisor saying what he says. I will try to provide examples showing both Searle's bolder proposal and Armstrong's modest one to be mistaken.

Suppose that a long-simmering feud between two Appalachian mountain families has erupted in violence. The patriarch of one of the families has been mortally wounded—say shot in the chest—and the wound prevents him from speaking even though he is conscious. Before his death his sons gather around him and the eldest says 'Pa, we promise you this, we will avenge your death'. Although the speaker believes that his father wants his death avenged, and has good reason for believing this, he is mistaken. The father has, in his last moments, come to the realization that continuing the blood feud will only bring further hardship to members of his family. He does not want his death avenged, and tries to say as much. But his wound prevents him from speaking before he dies, and his sons continue to believe that their father wanted his death avenged.

<sup>1</sup> I am indebted to F. Michael, N. Richards and a referee for helpful comments.

<sup>2</sup> 'Meaning and Communication', *The Philosophical Review*, LXXX (October, 1971), p. 446.

Has a promise been made? Certainly, the sons think one has; as they plot revenge they speak repeatedly of 'our promise to Pa'. And when they later succeed in murdering a member of the clan with which they are feuding they claim, with much satisfaction, that they have kept their promise to their father. Promises cannot be kept which are never made. And to my way of thinking there is no good reason to deny that a promise has been made in such a case, or for saying that the promise which was made was not made sincerely, or was somehow 'defective'. Although I think that most of us know what it means to call a part of an automobile or wrist watch defective, I don't think it is clear what it means to call a promise defective. Searle tells us that 'A promise is defective if the thing promised is something the promisee does not want done; and it is further defective if the promisor does not believe the promisee wants it done' (p. 58). But why, exactly, is this? The only reason Searle offers is that 'a non-defective promise must be intended as a promise and not as a threat or warning'. But the son who tells his dying father that he promises him his death will be avenged surely does not intend to threaten his father, or to warn him. Perhaps he does intend to reassure him, or to comfort him; but is there any reason he cannot both be doing these things and at the same time making a promise (or doing the former by doing the latter)? And even if the son were threatening the father it isn't entirely clear that he would not be promising him something in the process. A newly-wed husband whose bride promises never to leave his side may well feel threatened—the bride even may intend it as a threat; but it is not obvious that she has not made a promise in the process.

Perhaps it will be allowed that the son has, in the circumstances in question, promised sincerely and non-defectively (whatever this means) to avenge his father. But it may be added that had the son not believed (falsely as it happens) that his father desired to be avenged the son would not have been promising when he told his father that he promised to avenge his death. This amounts to arguing that Armstrong's condition can be salvaged even if Searle's proposal cannot.

Armstrong's more modest condition does seem more plausible, and less vulnerable to counter-example, than Searle's. But I am not convinced that Armstrong's proposal is correct. There are a variety of cases which could (and should) be discussed, but I will be concerned here only with this one: Suppose that I have a wealthy friend who is guilt-ridden concerning his many material possessions. He feels positively relieved when he manages to lose or destroy or give away or have stolen something he owns. And he is positively inconsolable when someone thoughtlessly presents him with a gift. But he cannot bring himself to give all of his possessions away, or simply to walk off and leave everything behind. Moreover, he is terribly embarrassed by his guilt-

feelings and does everything he can to conceal them from others. One day an acquaintance of his, Max, borrows a book from him, telling him that he promises to return it within a week. Max intends to return the book despite the fact that he believes (correctly, as it happens) that the person to whom his (putative) promise is made would prefer him not to return it. And Max believes himself to be morally obliged to return the book as a result of having said that he promises to do so.

This case satisfies two conditions which Armstrong places upon sincerely made promises: (1) that the person who makes the promise 'firmly intends' to do the thing promised, and (2) that he believes that, as a result of saying that he promises to perform the action in question, he is morally obliged to do so (pp. 445-446). But according to Armstrong, Max has not promised to return the book. He has not promised because he does not believe that the person who loaned him the book wants it returned. However, suppose that the sequel to our story goes this way: Max absent-mindedly keeps the book for several months. And I, knowing the facts of the case, accuse Max of having broken his promise. If Armstrong were right, the following could be said in Max's defence: 'Max cannot have *broken* a promise he never made; and he did not promise to return the book, since the person who loaned it to him did not want it returned and Max did not believe that he wanted it returned.' Surely this won't wash. We would not accept this defence, and we would be right not to. In saying what he says when he borrows the book, Max has caused people to have certain expectations. Even if its owner didn't want the book returned, he may well have been led to expect that it would be. He may have informed someone else, who needs the book to complete an important research project, that it would be available within a certain period of time. Or the book's owner may have postponed a trip he was looking forward to taking because he didn't want to put Max in the position of not being able to return the book to him by the appointed date. By not keeping his word Max may have let a number of people down—may have frustrated expectations which his own words inspired. Perhaps the implications of this point become somewhat clearer when the details of our story are altered, so that the book's owner is merely indifferent to its return. Surely his indifference, weighed against the possibility of the sorts of inconvenience just mentioned, does not defeat the claim that Max promised to return the book. Nor does the fact that Max believed the owner indifferent to the book's return serve to exonerate Max from the charge of having broken a promise. If this is right, believing the owner to want the book returned cannot be a genuine condition of Max's having made a promise.

Intuitively, perhaps, it seems correct to say that Max has promised to return the book; but our intuitions often play us false. Is there reason to believe they have done so in this case? Armstrong offers an interesting

reason in support of the claim that sincere promising can occur only where the promisor believes that the promisee wants the promise to be fulfilled. His argument turns upon a third alleged condition for making a promise sincerely: (3) In 'sincere promises the speaker's believing he is morally obligated to do the thing promised *depends* on his belief that the audience would prefer fulfillment to non-fulfillment' (p. 446). Where a person who says that he promises to do something does not believe his audience to prefer the performance of the action he says he promises to perform, and does believe he is morally obliged to perform this action, his belief that he is obliged presumably does not depend upon his belief that his audience prefers fulfillment. (In cases in which he does not believe himself morally obliged, he is not, given condition (2), promising sincerely.) And if the former belief does not depend, in some sense, upon the latter, condition (3) dictates that a sincere promise has not been made. Obviously this entails that Max's case fails to provide the sort of example we were looking for. Max's belief that he was morally obliged to return the book in no way depended upon his belief that the person who loaned it wanted it returned, since Max simply did not hold this last belief. So if (3) poses a genuine condition for sincerely-made promises, Max did not make such a promise, after all.

However, there seem to be good reasons for rejecting (3). People often believe that they are morally obliged to perform actions which they have never promised to perform. And some of these people sometimes promise to perform such actions. In such cases, the speaker's belief that he is morally obliged to perform the action which he promises to perform does not depend upon a further belief that his audience prefers that the action in question be performed. The speaker would believe himself morally obliged to perform this action even if he did not believe that his audience wanted it performed. We cannot make the blanket assessment, as (3) dictates, that those who are so situated never make sincere promises. Suppose that a man, Ralph, believes himself morally obliged to care for his aged and infirm father during the last years of the latter's life. Ralph's father periodically fears that his son will leave him to fend for himself. One day, to calm his fears, Ralph tells his father that he promises not to leave him on his own. In this case, Ralph believes (correctly) that his father wants him to fulfil his promise. However, Ralph's belief that he is morally obliged to provide continued help to his father does not depend upon his belief that his father wants him to do just that. As he grows older, Ralph's father finally may insist, sincerely, that he no longer wants Ralph's help. And Ralph may continue to believe himself obliged to help his father, even though he no longer believes that his father wants his help. Ralph's belief that he is morally obliged to help isn't dependent upon his belief that his father wants his help; but this hardly entitles us to conclude

## ON THE INTRINSIC WRONGNESS OF KILLING INNOCENT PEOPLE

By LEONARD GEDDES

IN what follows I intend to show two things. First, the principle 'It would always be wrong to kill an innocent human being, whatever the consequences of not doing so' does not apply (contrary to what some have thought) to all cases of the voluntary killing of the innocent. Second, this principle does not entail (contrary to what some have thought) anything about what it is impermissible to do.

### I

In a well known article Mr. Bennett says that the Principle of Double Effect is presupposed by the above principle.<sup>1</sup> I think that he is right in this. But if he is right, then something else he wants to say is false, namely that people who accept this principle concerning killing must claim, of a certain kind of obstetrical example, that it would be wrong for a surgeon to kill an unborn child by crushing its head so that the mother may live. The reason that Bennett is mistaken in this is that the Principle of Double Effect applies only to those voluntary actions that are intentional; although Bennett's obstetrical example involves the voluntary killing of the innocent, it does not involve the intentional killing of the innocent, which is the only sort of case to which the principle concerning killing can apply if it presupposes the Principle of Double Effect.

In Bennett's obstetrical example the surgeon has two mutually exclusive choices: either save the unborn child, where the only way to do this is to do things that will certainly mean that the mother will die, or save the mother, where the only way to do this is to crush the unborn child's head. Bennett argues that both alternatives have the same moral status intrinsically, even though in one of them the physician does actually kill a human being, thus the principle is indecisive in this example. His argument, which I find persuasive, is this:

Moral conclusions may be supported by facts (a) about what is expected, but in the example each upshot is confidently expected; (b) about what is inevitable, but in the example each upshot is inevitable; or (c) about what is ultimately aimed at, but in the example neither upshot is aimed at (p. 217).

Hence one must appeal to possible consequences in determining what to do.

<sup>1</sup> Jonathan Bennett, 'Whatever the Consequences', *ANALYSIS*, 26 (1966), 83-102. This article is reprinted in Judith J. Thompson and Gerald Dworkin (eds.), *Ethics* (New York, 1968), pp. 211-236, to which all references will be made.

If this is true, it follows, as Bennett implies, that the principle fails to apply decisively to every case of voluntarily killing the innocent. Since Bennett thinks the principle is meant to have such an application, he goes so far as to claim that this principle 'is in the last resort on a par with "It would always be wrong to shout, whatever the consequences of not doing so" or "It would always be wrong to leave a bucket in a hall-way, whatever *etc.*"' (p. 213). So anyone who held to this principle, 'if he was sincere and in his right mind, could safely be assumed to have delivered himself over to a moral authority and to have opted out of moral thinking altogether' (*loc. cit.*).

Miss G. E. M. Anscombe has done a great deal to clarify the importance of understanding that intentional actions only are referred to in absolute principles of the sort Bennett is attempting to discredit.<sup>1</sup> (Ironically, she is one of his main targets.) According to her, it is of the essence of the Principle of Double Effect to distinguish between the intended and the merely foreseen consequences of a voluntary action.<sup>2</sup> To be sure, the surgeon kills the unborn infant, but merely to say this does not help us see if it is intentional, i.e., the voluntary killing of a human being, either for its own sake or as a means to some end.<sup>3</sup> Clearly the surgeon does not kill the infant for the sake of killing it, as Bennett himself admits. So if Bennett's example is to discredit the principle against killing, the surgeon must kill the unborn child as a means to his end of saving the life of the mother.

At first glance it might seem that this is true, but a careful consideration of Bennett's example will show that it is not. A clear-cut case of choosing the killing of a person as a means to an end can be found in a familiar pragmatic justification for capital punishment. This justification contends that the execution of certain criminals will have certain desirable effects on society that imprisoning them for life will not have. But in the obstetrical example the *killing* of the infant cannot plausibly be thought of as having any effects desired by the surgeon; the death of the unborn child is not responsible, in any way, for the fact that the mother is now alive. The surgeon must remove the child from the mother's womb; the dimensions of the child are such that if the surgeon attempts to remove it without changing these dimensions the mother will surely die. He therefore alters these dimensions in certain ways. A necessary but quite unneeded and unwanted consequence of this procedure is that the child

<sup>1</sup> In, among other places, *Intention*, Oxford, 1957; 'Modern Moral Philosophy', *Philosophy*, 33 (1958), 1-19, reprinted in Thompson and Dworkin; 'War and Murder', in Walter Stein (ed.) *Nuclear Weapons: A Catholic Response* (New York, 1961), pp. 45-62, reprinted in Richard A. Wasserstrom (ed.), *War and Morality* (Belmont, 1970), pp. 42-53, to which all references will be made; and 'The Two Kinds of Error in Action', *Journal of Philosophy*, 60 (1963), 393-401, reprinted in Thompson and Dworkin.

<sup>2</sup> 'War and Murder', p. 50.

<sup>3</sup> As Miss Anscombe puts it: 'It is nonsense to pretend that you do not intend to do what is the means you take to your chosen end,' *ibid.*, p. 51.



dies. Clearly the death of the child does not enter into consideration as a means to anything. So, in the relevant sense, the killing of the child was not intended by the surgeon, either as an end in itself or as a means to an end. Hence it is a mistake to think that the principle concerning the killing of the innocent applies to the sort of killing found in Bennett's obstetrical example.

This argument does not imply, I should hastily add, the evil idea that as long as you do not intend to kill an innocent person you are never to be held responsible for or blamed for his death. To forestall this suggestion two additions to my argument are necessary. First, I think that the normal conscience will grant (the law certainly does) that you are responsible for non-intentional, voluntary actions if it was both necessary and possible for you to avoid them.<sup>1</sup> Second, I think that the normal conscience will also grant that a person may, generally speaking, be held responsible for the unintended evil consequences of his evil deeds, but is not to take credit for their good consequences; and conversely, given that the first qualification is not relevant, a person may not be held responsible for the unintended evil consequences of his good actions, even though he may get credit for their unintended good consequences. If these two points are added as qualifications to my argument, I think that it will not suggest anything evil.

I scarcely need to add here that I am not trying to provide a justification for the surgeon's killing the child. Whether it is necessary for the surgeon in a particular case to sacrifice the mother or the child will, as Bennett suggests, have to be decided on the details of the particular case. The necessity here will have to do with judgments about the beneficial or harmful consequences of each possible course of action, conditioned, of course, by considerations regarding rights and relevant prior obligations. As Miss Anscombe herself says, even if you hold to the idea that there are certain actions that are intrinsically wrong, 'there is a large area where what is just is determined partly by a prudent weighing up of consequences'.<sup>2</sup>

## II

I now turn to my second thesis. It has been thought by many philosophers, Bennett among them, that certain so-called "fantastic" examples can serve to discredit the principle 'It would always be wrong to kill an innocent human being, whatever the consequences of not doing so'. But, as we shall see, these examples could serve to discredit the principle only if the following two characteristically utilitarian principles are acceptable:

<sup>1</sup> Anscombe, 'The Two Kinds of Error in Action', Thompson and Dworkin, pp. 287-288.

<sup>2</sup> 'War and Murder', p. 50.

A. If something is the morally best thing to do under the circumstances, then it is right.

B. The morally best thing to do under the circumstances could never be something wrong.

I shall argue that these principles are not acceptable and that therefore the principle concerning the killing of the innocent cannot entail anything about what it is impermissible to do. Consequently, examples that turn on agonizing decisions about what to *do* cannot, on that account at least, pose a threat to one who accepts a principle which tells him only what is always *wrong*.

Miss Anscombe rightly charges W. D. Ross with being no better than the utilitarians when he claims:

The interests of the society may sometimes be so deeply involved as to make it right to punish an innocent man 'that the whole nation perish not'. But then the *prima facie* duty of consulting the general interest has proved more obligatory than the perfectly distinct *prima facie* duty of respecting the rights of those who have respected the rights of others.<sup>1</sup>

It is reasonable to assume that Ross's words presuppose principle A, since the determination of what *prima facie* duty is most obligatory (i.e., what is right *sans phrase*) will rest upon the discovery of the morally best thing to do under the circumstances. Let us say that execution is a possible sort of "punishment" in the kind of situation Ross is imagining, since this will give us the form of an example to which the principle regarding the killing of the innocent would be relevant: someone is intentionally going to kill an innocent man because the consequences of not doing so are unacceptable. Utilitarians survive on such examples.

As examples of this form are filled in by utilitarians it becomes clear that they are meant to present some proponent of the principle regarding the killing of the innocent with a choice between (1) doing what the normal conscience would reluctantly agree is, all things considered, the morally best thing to do under the hellish circumstances, and (2) holding to his principle. Anyone who chooses (2) over (1) would of course be charged with having opted out of moral thinking, of having given himself blindly over to some authority, etc. This line of attack clearly presupposes that you cannot embrace both (1) and (2). But you can embrace both if holding to your principle does not require you to believe that the act of intentionally killing the innocent is always impermissible. I shall now consider that possibility.

I believe that a little reflection will convince one that it is not false, self-contradictory or unintelligible to say, under certain circumstances:

<sup>1</sup> *The Right and the Good* (Oxford, 1930), p. 61. Anscombe's charge is made in 'Modern Moral Philosophy', Thompson and Dworkin, p. 196. She makes it clear here and elsewhere in this essay that she takes 'wrong' in absolute principles to mean 'impermissible under any circumstances'.

'There are times when a good man must do something he knows to be wrong', where these words imply that the man is morally justified (as opposed to being justified on grounds of expediency) in doing what he knows to be wrong. If 'wrong' meant 'not to be done under any circumstances' moral conflicts that we are all too painfully aware of would not be possible. I am not here referring to conflicts between one's passion or greed and his knowledge that he has no justification for doing something in a particular situation. This is an example of temptation, not moral conflict.

I am referring to the sort of case in which one does something he knows to be wrong, but which, upon painful consideration, he has correctly judged to be morally justified under the circumstances. For example, suppose you have to betray a strict confidence (where this is a very serious thing to do to the blameless friend who entrusted you with it) in order to ransom your child. The case is one in which, by letting out some confidential information about certain stock you make a quick one hundred thousand dollars and cause your friend to lose the chance of making that much. Suppose you are morally justified in doing this; this fact does not make your betrayal of your friend right. You know that in betraying his trust you have done something wrong.<sup>1</sup> It is for this reason that you ask him to forgive you.

I conclude, then, that both Miss Anscombe and the utilitarians, against whom she argues, have misunderstood the principle, 'It would always be wrong to kill an innocent human being, whatever the consequences of not doing so'. For (contra Anscombe) the morally best thing to do under certain circumstances might be something wrong in itself. From this we can see that if something is the morally best thing to do under certain circumstances we need not, on that ground, (contra the utilitarians) conclude that it was right. In regard to the sort of case Ross was imagining, we can say that perhaps, in some never-never land example, a leader entrusted with the lives and safety of his people might be morally justified in killing an innocent man 'that the nation perish not'. However, we must not say with Ross that killing that innocent man would be right, even in this imaginary hell. It would be terribly wrong, terribly unjust, and we know it. The principle 'It would always be wrong to kill an innocent human being, whatever the consequences of not doing so', does capture something basic in the normal conscience.

<sup>1</sup> I think that it is a mistake to suppose, as some have, that the person who ransoms his child and admits that he had to do something wrong in order to achieve this, must feel that he has failed to live up to some *standard* expressed in the statement, 'It is always wrong to betray the confidence of a blameless friend'. I think that this principle obviously expresses something a person can think true or false. But standards are neither true nor false. For the same reason this and other absolute principles cannot be thought of as rules, either.

## ON THE COHERENCE OF ACT-UTILITARIANISM

By LARS BERGSTRÖM

PROFESSOR Hector-Neri Castañeda has drawn attention to 'the problem of formulating a coherent act-utilitarianism' (see his note 'A Problem for Utilitarianism', *ANALYSIS* 28.4 (1968), pp. 141-142, and also his longer article 'Ought, Value, and Utilitarianism', *American Philosophical Quarterly*, Vol. 6 (1969), pp. 257-275). This problem, or at least one version of it, can be roughly indicated as follows: If an agent X ought to do a conjunctive act A & B in circumstance C, then, according to a well-known principle of deontic logic, X ought to do A in C. But this principle does not seem to be satisfied by the usual versions of act-utilitarianism. For X's doing A & B in C may very well be optimistic even though X's doing A in C is not.

In a reply to Castañeda ('Utilitarianism and Deontic Logic', *ANALYSIS* 29.2 (1968), pp. 43-44) I therefore suggested the following formulation of act-utilitarianism:

(U) X is obliged to do A in circumstances C if, and only if, some particular action which consists in X's doing A at some specific time in circumstances C satisfies one of the following conditions: (i) it is a member of some relevant alternative-set and its consequences are better than those of any of its alternatives [in this set], or (ii) it is a component [conjunct] of some compound [conjunctive] action which satisfies (i) above, or (iii) it is a compound [conjunctive] action each of whose components [conjuncts] satisfies either (i) or (ii) above.

This, I thought, would take care of Castañeda's problem. Of course, a crucial notion in (U) is that of a 'relevant alternative-set'. I have discussed this notion at some length in the second chapter of my book *The Alternatives and Consequences of Actions* (Stockholm 1966) and also in 'Utilitarianism and Alternative Actions', *Notas*, Vol. 7 (1971), pp. 237-252.

However, in a recent note ('On the Problem of Formulating a Coherent Act-Utilitarianism', *ANALYSIS* 32.4 (1972), pp. 118-124) Castañeda claims that I have not yet solved his problem. He seems to have nothing against clauses (ii) and (iii) in (U), but he is unwilling to accept (U) as a whole. He says that clause (i) 'is beset with some difficulties'.

The first difficulty is that 'an action is made obligatory by *some* relevant alternative-set. This is dangerous, for there is no guarantee that different alternative-sets will yield compatible results' (*ibid.*, pp. 123-124). But this is surely beside the point. The question is rather

whether different *relevant* alternative-sets will yield compatible results. I can see no reason why they should not do so, but I shall return to this question below.

Secondly, however, Castañeda claims that I have not solved the problem of determining *the* relevant alternative-set in a given case. 'Bergström discusses nine criteria for determining the relevant alternative-set. But those criteria are just necessary conditions for being the relevant alternative-sets. Bergström does not furnish a single sufficient condition to complete his formulation of act-utilitarianism' (*ibid.*, p. 124). To a certain extent I plead guilty to this charge. But even if my criteria do not always determine a *unique* alternative-set as *the* relevant one in a given case, they may still be *jointly* sufficient from the point of view of (U): it may still be held that every alternative-set which satisfies my criteria is *a* relevant one. And again, the important question seems to be whether different alternative-sets which satisfy my criteria for being relevant can ever yield incompatible results.

Thirdly, in 'Ought, Value, and Utilitarianism' Castañeda introduces a distinction between ' $\alpha$ -alternativeness' and ' $\mu$ -alternativeness'. He now says: 'I show that the former is much less satisfactory than the latter.... Bergström's alternativeness includes, and is equivalent to, my  $\alpha$ -alternativeness. Thus, the arguments given in my paper apply to clause (i) in Bergström's formulation of act-utilitarianism' (*ibid.*). Let us consider this charge, which—according to Castañeda—is the most serious one.

In 'Ought, Value, and Utilitarianism'  $\alpha$ -alternativeness is explained as follows. We start with a given agent in a given situation of choice. We suppose that 'the primary, non-analyzed actions that he is concerned with are  $A_1, \dots, A_{n-1}, A_n$ ' (p. 260). Every  $n$ -termed conjunction of these primary actions or their negations is called a 'basic utilitarian action'. Basic utilitarian actions as well as disjunctions of basic utilitarian actions are called 'utilitarian actions'. 'Now, a set  $\alpha_k(A_1 \dots A_n)$  of utilitarian actions all of whose members are incompatible by pairs and jointly exhaustive of the open possibilities of action will be called a set of  $\alpha$ - or  $\alpha_k$ -alternatives' (*ibid.*, p. 261).

As an illustration Castañeda invites us to consider a case with only two primary actions, A and B. These primary actions determine, he says, the following 'possible sets of open  $\alpha$ -alternatives':

$$\begin{aligned} \alpha_1(AB) &= \{A \ \& \ B, A \ \& \ \sim B, \sim A \ \& \ B, \sim A \ \& \ \sim B\} \\ \alpha_2(AB) &= \{A \ \& \ B, A \ \& \ \sim B, \sim A \ \& \ \sim B\} \\ \alpha_3(AB) &= \{A \ \& \ B \vee \sim A \ \& \ \sim B, \sim A \ \& \ B\} \\ \alpha_4(AB) &= \{A \ \& \ B \vee \sim A \ \& \ \sim B, A \ \& \ \sim B \vee \sim A \ \& \ \sim B\}, \text{ etc.} \\ & \quad (\textit{ibid.}) \end{aligned}$$

The notion of a set of  $\alpha$ -alternatives is *not* the same as, or even equivalent to, my notion of an 'alternative-set', nor is it equivalent to

my notion of a 'relevant alternative-set'. I require that every action in an alternative-set be performable (or 'open', in Castañeda's terminology), but every action in a set of  $\alpha$ -alternatives need not be performable (open). For example, if  $\alpha_2(AB)$  above is a set of  $\alpha$ -alternatives, then  $\sim A \ \& \ B$  is not open to the agent in the given situation, but  $\alpha_1(AB)$  would still be a set of  $\alpha$ -alternatives according to Castañeda's definition.

On the other hand, if we suppose that a set of *open*  $\alpha$ -alternatives is a set of  $\alpha$ -alternatives every member of which is open to the agent in the circumstances, then a set of open  $\alpha$ -alternatives seems to be much the same as an 'alternative-set' in my sense. But it should then be noted that at most one of the four sets in Castañeda's example can actually be a set of *open*  $\alpha$ -alternatives in the case in question. (For example, if  $\alpha_2(AB)$  is such a set, then neither  $\alpha_1(AB)$  nor  $\alpha_3(AB)$  is such a set—since they contain a member, viz.  $\sim A \ \& \ B$ , which is not open—and neither is  $\alpha_4(AB)$  such a set—since its members are not incompatible when  $\sim A \ \& \ \sim B$  is open.)

There may, however, be more than one set of open  $\alpha$ -alternatives in a given case. For example, if  $\alpha_2(AB)$  is such a set, then so is

$$\alpha_5(AB) = \{ A \ \& \ B \vee A \ \& \ \sim B, \ \sim A \ \& \ \sim B \}.$$

Suppose that the consequences of  $A \ \& \ B$  are equally good as those of  $A \ \& \ \sim B$  but better than those of  $\sim A \ \& \ \sim B$ . ('The consequences of  $A \ \& \ B$ ' is short for 'the consequences of the agent's doing  $A \ \& \ B$  in the given circumstances', and so on.) I believe that Castañeda is referring to cases of this kind when he says: 'In some situations we ought<sub>i</sub> to do ( $A_1 \ \& \ A_2 \vee A_1 \ \& \ \sim A_2$ ), even though it is not the case that we ought<sub>i</sub> to do one disjunct rather than the other' (*ibid.*). Anyway, this exemplifies a case where a 'disjunctive' act has to be taken into account. We cannot very well regard  $\alpha_2(AB)$  as *the* alternative-set here, for the action which ought to be done, viz.  $A \ \& \ B \vee A \ \& \ \sim B$ , or simply  $A$ , is not a member of this set. Notice, however, that so far this presents no difficulty for a criterion like (U)—even if we substitute 'set of open  $\alpha$ -alternatives' or 'alternative-set' for 'relevant alternative-set'.

But now suppose instead that the consequences of  $A \ \& \ B$  are better than those of  $A \ \& \ \sim B$  and also better than those of  $\sim A \ \& \ \sim B$ , but that the consequences of  $\sim A \ \& \ \sim B$  are better than those of  $A \ \& \ B \vee A \ \& \ \sim B$ . Then  $\alpha_2(AB)$  and  $\alpha_5(AB)$  'yield incompatible results', so to speak. (We are assuming throughout that  $\alpha_2(AB)$  is a set of open  $\alpha$ -alternatives.) More precisely, if 'alternative-set' is substituted for 'relevant alternative-set' in (U), it turns out that, according to (U), both  $A \ \& \ B$  and  $\sim A \ \& \ \sim B$  ought to be done by the agent in the circumstances. This is unacceptable since these actions are incompatible. The same difficulty arises if we use Castañeda's utilitarian principle for  $\alpha$ -alternatives, which he formulates as follows:

( $O_\alpha$ ) X ought<sub>i</sub> to do a utilitarian action A in circumstances C, if and only if: there is a set  $\alpha_n(A_1 \dots A_n)$  of  $\alpha$ -alternatives open to X in C such that A belongs to  $\alpha_n(A_1 \dots A_n)$  and for any member Z of this set different from A the value of kind  $K_i$  that X would bring about by doing A is greater than the value of kind  $K_i$  that X would bring about by doing Z. (*Ibid.*)

Castañeda may be referring to difficulties of this kind when he says that he has shown that  $\alpha$ -alternativeness 'is much less satisfactory than'  $\mu$ -alternativeness. This is only a guess, however, for his own formulations are rather puzzling. He says that 'it is still open whether or not by ( $O_\alpha$ ) an action is obligatory<sub>i</sub> because of its membership in a given set of  $\alpha$ -alternatives, and non-obligatory<sub>i</sub> because of its membership in another set of  $\alpha$ -alternatives' (*ibid.*). As an example of two such sets he mentions  $\alpha_3(AB)$  and  $\alpha_4(AB)$  (*ibid.*). And later he says: 'In the case of  $\mu$ -alternatives we do not have to worry about actions being both obligatory<sub>i</sub> and non-obligatory<sub>i</sub> on account of their membership in different sets of alternatives' (*ibid.*, p. 262). But, in the first place,  $\alpha_3(AB)$  and  $\alpha_4(AB)$  can hardly present any problem for ( $O_\alpha$ ) since they cannot both be sets of *open*  $\alpha$ -alternatives in a given case. And secondly, if an action is both obligatory<sub>i</sub> and non-obligatory<sub>i</sub> according to ( $O_\alpha$ ), this is hardly because of its membership in different sets of open  $\alpha$ -alternatives, but rather because of certain principles of deontic logic in conjunction with the fact that two incompatible actions are both obligatory<sub>i</sub> according to ( $O_\alpha$ ). For example, in the last case described above it might be argued that A & B is both obligatory<sub>i</sub> and non-obligatory<sub>i</sub> according to ( $O_\alpha$ ). But this is not because it is a member of different sets of open  $\alpha$ -alternatives. So far, we have only assumed that it is a member of one such set, viz.  $\alpha_2(AB)$ .

The kind of difficulty exemplified here remains, however, even if Castañeda's argument against  $\alpha$ -alternativeness is inconclusive. Hence, (U) cannot be accepted if 'alternative-set' or 'set of open  $\alpha$ -alternatives' is substituted for 'relevant alternative-set'. (Neither, of course, can ( $O_\alpha$ ) be accepted.) But things seem to be different if we take (U) as it stands and concentrate upon *relevant* alternative-sets.

Let us go back to the two examples outlined above. Of course, these are rather artificial. We have assumed that the agent is only 'concerned' with two primary actions, A and B, but it is surely unrealistic to assume that these are in fact the only primary actions open to him in the circumstances. In practice, other primary actions may also have to be taken into account when we try to decide whether a given alternative-set is relevant. We may even have to take account of actions which will be open to the agent in certain future circumstances. (See the discussion between Dag Prawitz and myself in *Theoria*, Vol. 34 (1968), pp. 76-84 and pp. 163-170, and Vol. 36 (1970), pp. 116-126.)

However, in the first example (where the consequences of  $A \ \& \ B$  are equally good as those of  $A \ \& \ \sim B$  but better than those of  $\sim A \ \& \ \sim B$ ), it seems clear that  $\alpha_2(AB)$  is *not* a relevant alternative-set according to the criteria suggested in my book (*The Alternatives and Consequences of Actions*, pp. 51–53). But so far as we can tell from the example, it seems that  $\alpha_5(AB)$  is a relevant alternative-set in this case. If so, we may conclude that  $A \ \& \ B \vee A \ \& \ \sim B$  ought to be done according to (U). This seems quite acceptable.

In the second example (where the consequences of  $A \ \& \ B$  are better than those of  $A \ \& \ \sim B$  and also better than those of  $\sim A \ \& \ \sim B$ , but where the consequences of  $\sim A \ \& \ \sim B$  are better than those of  $A \ \& \ B \vee A \ \& \ \sim B$ ), *neither*  $\alpha_2(AB)$  *nor*  $\alpha_5(AB)$  is relevant according to my criteria. On the other hand, it seems that

$$\alpha_6(AB) = \{A \ \& \ B, A \ \& \ \sim B \vee \sim A \ \& \ \sim B\}$$

is a relevant alternative-set in this case. If so, it turns out that  $A \ \& \ B$ —as well as  $A$  and  $B$ —ought to be done according to (U). For all we know, this is also as it should be, and no inconsistencies seem to arise.

Of course, by arguing that my criteria behave as they should in certain problematic cases I cannot claim to have *shown* that (U) is a coherent formulation of act-utilitarianism. But neither, it seems, has Castañeda shown that it is not. He may be right in holding that I have not yet solved his problem, but it seems to me that (U) has at any rate survived his attacks so far.

*Stockholm University*

## REASONS FOR ACTING VERSUS REASONS FOR BELIEVING

By GARY A. WEDEKING

**I**N a recent paper (ANALYSIS 32.1) H.-N. Castañeda discusses a distinction that I make (ANALYSIS 30.5) between reasons for believing and reasons for acting. He does not deny that such a distinction exists. Indeed, he claims to have 'insisted all along' on it. What he finds objectionable is my claim that while reasons for believing are closely related to possible inferences (in which those reasons constitute the grounds for the inference) so that they might reasonably be called 'inferential', reasons for acting are not. In this paper I wish to defend this claim against Castañeda's objection and to refute the account of reasons for acting in terms of inference which he offers in its place.



In favour of my claim, I pointed out that while

3. Since  $p$ , it is the case that  $q$

is closely related to such statements as

6. From the fact that  $p$  it follows that  $q$ ,

there is no similar relationship between a statement in which the reason given is a reason for acting, viz:

4. Since  $p$ , (I promise) I shall do  $x$

5. Since  $p$ , do  $x$

and the corresponding statements (respectively)

8. From the fact that  $p$  it follows that I will do  $x$

9. From the fact that  $p$  it follows that you will do  $x$ .

Where I have gone wrong, according to Castañeda, is in my formulations of 8 and 9. He develops three alternative formulations of 9. Since the first two are very similar and the third lacks even an apparent relationship to 5 (and would furthermore entrap us in the quagmire of deontic logic), we need only examine the first:

- 9.1. 'John, go home' follows from, and was inferred by Smith from 'John, go home and study'.

One's attitude to 9.1 will obviously depend on whether one holds that there is anything which could properly be called command or imperative inference. But quite apart from that matter, it should be noted that in order to take 9.1 as the counterpart of 5, the sentence letters (' $p$ ' and ' $q$ ') of 5 must be interpreted as follows:

- 5.1. Since, John, go home and study, John, go home.

Since both Castañeda and I now agree that this is nonsense<sup>1</sup>, one wonders how Castañeda could have selected a reading of ' $p$ ' which forces such a reading of 5. This can only be explained, I believe, by the fact that a sensible reading of 5 would not yield any suitable reformulation of 9. When I formulated 5 with schematic sentence letters I assumed that the context would make it clear that I intended it not as another logic-book example of imperative entailment but rather as an instance of the everyday practice of giving orders, or, more likely, requests, together with reasons for carrying them out. What I had in mind were utterances like

<sup>1</sup> (Castañeda, pp. 13-14.) By implication Castañeda must also concede that my refutation of his argument that there are command arguments goes through. The point that he develops at length in his reply, that this claim does not establish that there is no such thing as command inference, is, of course, a red herring. I explicitly pointed out that my argument was only *prima facie* evidence in favour of this latter claim.

- 5.2. Since you have an examination tomorrow, John, go home and study.

Now according to Castañeda, what was wrong with my original formulation of 9 was that what it asserted to follow from the fact that  $p$  was that *you will do  $x$*  rather than the *command or decision to do  $x$* . So a reformulation of 9 (corresponding to 5.2) which should satisfy Castañeda would be

- 9.4. The command or decision for John to go home and study follows from the fact that John has an examination tomorrow.

But this seems just as implausible as 9 itself on the face of it. Could even the most committed of command logicians be so bold as to declare such an "inference" valid? In what sense could a command or decision be said to "follow" from such a consideration? Is John's girl friend logically compelled to order him out when he informs her of the impending examination? Or must John, in order to avoid self-contradiction, decide to study when he remembers the exam? Surely this is merely a parody of practical reasoning. If John yields to temptation he is not guilty of an error in logic. Nor is weakness of will all that is relevant. As I pointed out in my earlier paper, reasons for doing something may be contrasted with reasons against doing it. Even the best of reasons in favour may in some circumstances be judged by John—and correctly judged—to be outweighed by reasons against doing it.

It is important to note the fact that there may be reasons against, as well as reasons for, an action in connection with Castañeda's attempt to account for the distinction between reasons for acting and reasons for believing. A reason for doing, he states, is a member of a premise set 'that is taken to imply an order, command, request, entreaty, or a decision' (p. 150). In discussing this thesis we shall ignore its reference to inferred orders, commands, requests and entreaties. This is justified by the fact that writers such as R. M. Hare and Castañeda, who find it necessary to explain their concept of implied commands, etc., in terms of 'acceptance' (Hare) or 'endorsement' (Castañeda, p. 16) by an addressee of the command (imperative, etc.), implicitly collapse their notion of an implied *command* into that of an implied *decision*. I believe it can be shown that essential reference to speech acts such as commands and orders drops out from these theories altogether, but limited space prevents me from arguing that here.

The existence of reasons against actions proves untenable the idea that there could be a complete set of reasons implying a decision. For, unlike a conclusion which is implied by true premises, the result of a practical argument is, in Hart's terminology, defeasible. It makes perfect sense to say 'that  $p$  and that  $q$  are good reasons for doing  $x$ , but since  $r$ ,

you should not do  $x$ ', regardless of the scope of the reasons stated in ' $p$ ' and ' $q$ '. But one cannot say 'The fact that  $p$  implies that  $q$ , but since  $r$ , not- $q$ .' If 'not- $q$ ' is true then either ' $p$ ' does not state a fact or that fact does not imply that  $q$ . If we want to know whether it is the case that  $p$ , all we need to know is a fact or set of facts which implies it. But to make an informed decision, it is not sufficient to be apprised merely of reasons for the action, even though those reasons would *ceteris paribus* constitute overwhelmingly good reasons for the action. No matter how good the reasons are, there is always the possibility of discovering contrary reasons that outweigh them.

It must be conceded that the above argument shows Castañeda's thesis to be wrong only in detail rather than in broad outline. For the concept of inference in general is unlike that of implication with respect to the defeasibility of conclusions. That is, one may infer that  $p$  from the fact that  $q$ , nevertheless conclude that not- $p$  when other facts are pointed out or discovered. A proposition (finding, hypothesis, theory) for which we have evidence is similarly defeasible by discovering contrary evidence. (This shows, incidentally, that it is a conceptual mistake to speak of "inductive implication".) So it appears that Castañeda's thesis could be reconstructed in terms of inference and/or evidence instead of implication. Thus it would read 'A reason for acting is a member of a set from which a decision may be (reasonably?) inferred (or which constitutes evidence for the decision)'.

The main defect in this thesis resides in the very idea of giving evidence for or inferring a decision rather than in any detail of formulation. Castañeda does not tell us what it is about the decision that is supposed to be inferred (or implied, in his formulation). We understand what it is to *make* a decision, but what is it to *infer* one? We may infer a decision to be such and such, or give evidence that it is thus and so. But we would hardly know what to say if asked whether we could give evidence for or infer a certain decision, full stop. One might try to make sense of this by using 'inferring a decision from  $p$ ' idiosyncratically to *mean* deciding because of  $p$ . Although one suspects that this may be Castañeda's intention, such a notion of inference, requiring, as it would, definition in terms of 'deciding for a reason', would obviously be unhelpful in analyzing the concept of a reason for decision or action. It would thus be useless for Castañeda's purpose. If, on the other hand, we make use of the ordinary concept of inference, then we must ask what it is about the decision that is supposed to be inferred. Is there, then, a way in which this notion could be plausibly filled in? Does Castañeda have in mind, perhaps, that what we infer is that a decision was (or will be) made or that it is a good decision?

It is true that we are sometimes able to infer that a person will (or did) decide to do  $X$ , using as our evidence the fact that there exists a certain

reason or set of reasons for doing X. But such an inference is only possible when we know something of the actual motivation of the person making the decision. There are often reasons for acting which are not in fact acted upon, because, for example, the agent considers them outweighed by other reasons, or does not consider them good reasons, or simply fails to think of them. Furthermore, whatever an *observer* might do in trying to predict a decision, when the *agent* himself considers the reasons in the course of deliberation, *he* is not trying to infer or consider evidence for how he will decide. Predicting how a person (even oneself) will decide and deciding are, after all, rather different activities. You do not, for example, have to consider facts about your character and disposition in order to make a decision; you do generally have to consider such facts in order to predict how you will decide.

It is most plausible to construe the notion of inferring a decision as inferring that a decision is a good decision. An attempt to formulate Castañeda's thesis along these lines would have to overcome the *prima facie* difficulty that there are often reasons for doing something even though the decision so to act would not be a good one. There may be reasons for acting in a certain way, but better reasons for acting in some other, contrary way. But it must be conceded that there is a connection between good reasons and good decisions. The nature of this connection, however, vitiates any attempt to account for the former in terms of a relationship with the latter. For the connection is such that we could not understand the concept of a good decision unless we understood that of a good reason. A good decision to do something is one for which there are good reasons and no *manifestly* better reasons to do something else. When we discuss whether a decision was a good one, what we talk about *are* the reasons for and against it. To evaluate a decision *is* to evaluate the reasons upon which it is based. So it is not surprising that the existence of a good reason for acting is often evidence for the fact that the decision to act is, or would be, a good one. But to try to get this fact to bear the burden of a conceptual analysis of a reason for acting is obviously to put the cart before the horse. It is rather that this evidential relationship follows (albeit loosely) from the conceptual relationship, which is *not* evidential, between good decisions and good reasons. We conclude, therefore, that the prospects for an adequate analysis of the concept of a reason for acting on the lines suggested by Castañeda are dim indeed.

*University of British Columbia*

## MARTIN ON EXPLANATION AND CONFIRMATION

By BARRY GOWER

IN a recent study of qualitative confirmation, Brody<sup>1</sup> suggested that the following two conditions might prove acceptable:

- (1) If evidence E confirms  $H_1$  and  $H_1$  explains  $H_2$ , then E confirms  $H_2$ ;
- (2) If evidence E confirms  $H_1$  and  $H_2$  explains  $H_1$ , then E confirms  $H_2$ .

Martin<sup>2</sup> has argued that (2) is false on the grounds that there are cases where E confirms  $H_1$  and  $H_2$  explains  $H_1$ , but E disconfirms  $H_2$ , and also cases where E confirms  $H_1$  and  $H_2$ , but  $H_2$  does not explain  $H_1$ . Since condition (2) only specifies conditions which are sufficient for confirmation, it is difficult to see why cases of the second kind should be thought relevant to the acceptability of (2). Martin is mistaken in thinking that this condition can exclude confirmations which on intuitive grounds seem reasonable and I shall ignore the example he uses to indicate this possibility.

If there are any cases of the first kind, they would prove that certain confirmations allowed by (2) are counter-intuitive. Martin supports his claim that there are such cases by means of the following example:

$E_1$ : Mr. X lives in Chinatown and most people who live in Chinatown are law-abiding and most people who live in Chinatown are not fundamentalist Baptists.

This evidence confirms  $H_1$ :

$H_1$ : Mr. X is law-abiding.

Also, we can suppose  $H_1$  to be explained by  $H_2$  (here slightly abbreviated):

$H_2$ : Mr. X is a fundamentalist Baptist and most fundamentalist Baptists are law-abiding.

But, Martin says,  $E_1$  disconfirms  $H_2$  whereas according to (2) it should confirm  $H_2$ . Consequently (2) is false because it would force us to accept disconfirming evidence as confirming evidence.

Presumably the reason for saying that  $E_1$  disconfirms  $H_2$ , *i.e.* confirms  $\sim H_2$ , is that  $E_1$  confirms  $H_3$ :

$H_3$ : Mr. X is not a fundamentalist Baptist,

<sup>1</sup> B. A. Brody, 'Confirmation and Explanation', *Journal of Philosophy*, 1968, pp. 282-299.

<sup>2</sup> M. Martin, 'Confirmation and Explanation', *ANALYSIS*, 32.5, 1972, pp. 167-169.

and  $H_5$  entails  $\sim H_2$ . That is to say, the claim that  $E_1$  disconfirms  $H_2$  depends upon the acceptance of the special consequence condition:

(1') If evidence  $E$  confirms  $H_1$  and  $H_1$  entails  $H_2$ , then  $E$  confirms  $H_2$ .

Thus Martin's example shows, not that (2) is false, but that it is incompatible with (1'). Since  $H_5$  does not explain  $\sim H_2$ , the counter-example could be avoided by rejecting (1') in favour of (1).

In this connection, it is worth noticing that an example similar to Martin's can be produced to cast doubt on condition (1). For example, suppose that instead of  $E_1$  we have  $E_2$ :

$E_2$ : Mr. X lives in Chinatown and most people who live in Chinatown are not law-abiding and most people who live in Chinatown are fundamentalist Baptists.

This evidence appears to confirm  $H_2$ , and  $H_2$ , as before, explains  $H_1$ . But, contrary to (1),  $E_2$  disconfirms  $H_1$ . However, the grounds for claiming that  $E_2$  confirms  $H_2$  are that it confirms  $\sim H_5$  and that  $H_2$  entails  $\sim H_5$ . Thus, the example indicates, not that (1) is false, but that it is incompatible with the converse consequence condition:

(2') If evidence  $E$  confirms  $H_1$  and  $H_2$  entails  $H_1$ , then  $E$  confirms  $H_2$ .

Evidently, the use of either example to show that (1) or (2) is false rests on the assumption that certain other confirmation conditions, which Brody was concerned to replace, are true.

Nevertheless, it might still be urged, so far as Martin's example is concerned, that  $E_1$  disconfirms  $H_2$ . I take it that this amounts to the demand that, in order to satisfy our "intuitions", conditions (2) and (1') ought to be compatible. Whether Martin's example shows that this demand is impossible to satisfy depends upon what is to be understood by 'explanation' in the context of (2). In particular, it is not clear that, given  $E_1$ ,  $H_2$  can be used to explain  $H_1$  unless this means merely that  $H_2$ , if true, would make  $H_1$  highly likely. Suppose that instead of  $E_1$  we have  $E_3$ :

$E_3$ : Mr. X lives in Chinatown and most people who live in Chinatown are law-abiding and no people who live in Chinatown are fundamentalist Baptists.

Although  $E_3$ , like  $E_1$ , confirms  $H_1$  and disconfirms, by falsifying,  $H_2$ , we would not accept  $H_2$  as an explanation of  $H_1$  even though it would, if true, make  $H_1$  highly likely. In Martin's example  $E_1$  does not, admittedly, make  $H_2$  false but, in the absence of any further relevant evidence, it does make its truth highly improbable and, to that extent, unacceptable as an explanation of  $H_1$ . In general, the compatibility of (2) and (1') could be upheld despite Martin's example by insisting that if

our available evidence disconfirms a hypothesis then that hypothesis is not acceptable as an explanation of anything even though some of its deductive or inductive consequences may be confirmed. Whether or not such an insistence can be justified in terms of an adequate analysis of explanation, it remains true that Martin has failed to show that Brody's condition (2) is false.

*Durham University*

## SMART ON CONDITIONS OF IDENTITY

*By* LAWRENCE H. DAVIS

BRIAN Smart apparently believes<sup>1</sup> that identity of parts and spatio-temporal continuity of form are 'two independent, but defeasible, sufficient conditions of identity' of *equal weight*: either may be defeated by applicability of the other. For his judges decide in favour of Bombos only because they deem the first of these conditions inapplicable. But their argument makes sense only if the continuity of form condition has *greater weight*; and if so, their argument is superfluous.

The judges claimed that 'when a plank had been removed from X and replaced by a new plank, it was the new plank that was now a part of X. The old plank *had been* a part of X but was no longer' (p. 148). Therefore the ship Z formed out of the reassembled old planks could not be X by the identity of parts condition: Z had *no* parts in common with X, so the condition was inapplicable.

But why were the old planks no longer "parts of X" in the sense required for applicability of the condition? Just because they had been removed from X? Surely not, else the identity of parts condition could *never* be sufficient. As soon as a ship were decomposed into a heap of planks, those planks would cease being "parts of X", and so their reassembly could not restore the original ship. So the old planks must have lost their status as "parts of X" because each had been *replaced* as well as removed. In other words, each old plank ceased being a part of X in the relevant sense because X continued to exist, intact, without it. Continuity of the form of X was preserved throughout the time in which the old planks were being gathered and assembled into Z. But the "relevant sense" here is just the sense required for applicability of the identity of parts condition; so to say that the continuity of X's form destroys the status of the old planks is just a roundabout way of saying

<sup>1</sup> 'How to Reidentify the Ship of Theseus', Brian Smart, *ANALYSIS*, April 1972, pp.145-8.

that the continuity of form condition, where applicable, takes precedence over the identity of parts condition.

The claim that the conditions are of equal weight might be defended in the face of this argument by claiming that though sometimes continuity of form takes precedence, cases are imaginable also in which identity of parts would take precedence. Suppose for example that only Bombos is involved. He has the old planks of X removed and replaced one by one as before, and simply gathered into a pile over which he retains ownership. Next suppose that Y—the renovated X—is destroyed by fire, and *then* Bombos has the collected old planks assembled into a ship Z, which *he* insists on regarding as X. Z does not have any parts in common with Y, so if he is right, identity of Z's parts with X's parts here takes precedence over continuity of Y with X, and my reference to Y as 'the renovated X' was incorrect.

I have some sympathy for Bombos, but less for the view I now represent him as espousing. He would admit that until the fire it *seemed* that Y was identical with X; so he must insist that subsequent events can motivate *and warrant* retroactively revising our identity judgments. This might be tolerable in itself, but the kinds of motivation that it seems Bombos would accept as warranting the revision include such extraneous matters as his desire still to have his ship X, and the absence of any disturbing considerations such as ownership of the reassembled planks by Morion would be. Identity on this view is significantly a function of interests of the moment rather than deep-seated features of our conceptual scheme, and this rankles. It would mean among other things that we could not expect the 'existing rules of identity' to yield 'non-arbitrary and clear-cut decisions' in all important cases, a feature of this view which Smart also would apparently regard as unpalatable (cf. p. 148). We are left with the conclusion that spatio-temporal continuity of form (however this is to be spelled out) is the primary condition for identity of ships, and perhaps for all spatio-temporal objects.

*The Johns Hopkins University*



## RUSSELL'S INCOMPLETE SYMBOLS

By J. O. URMSON

I DO not wish to argue closely with Mr. R. K. Perkins, Jnr., regarding his article 'Urmson on Russell's Incomplete Symbols', which appeared in *ANALYSIS*, Vol. 32, No. 6. I entirely agree that Russell's official and technical account of incomplete symbols is without ontological implications and that Russell often stressed that to show that 'round square' or 'unicorn' was an incomplete symbol merely removed an argument in favour of the existence of round squares and unicorns and did not disprove their existence. So I should not have said that Russell held that 'to show that "X" is an incomplete symbol is tantamount to showing that there are no Xs'. But I am still willing to say that when Russell said that 'X' was an incomplete symbol his point was almost invariably to throw doubt on the reality of Xs.

Mr. Perkins plays down the ontological implications, which are indeed absent from the official definition of 'incomplete symbol', and warns us not to be misled by Russell's alternative expression, 'logical fiction', into believing that by 'fiction' Russell meant 'fiction'. But Russell is quite happy to say that 'numbers are logical fictions or symbolic conveniences' and that classes are 'symbolic conveniences'. This has not the ontologically aseptic character that Mr. Perkins ascribes to Russell's terminology; it is in fact presented by Russell as the technical way of making the point that we may legitimately say 'there are numbers' but that it does not follow that numbers are 'part of the furniture of the world'. For these quotations and other revealing remarks the reader may consult *My Philosophical Development* in the general area of page 234.

The essential point is that Russell so defined the expressions 'logical fiction' and 'incomplete symbol' that to say that Xs are logical fictions or that 'X' is an incomplete symbol is to say nothing about the reality of Xs; it is purely logical analysis. But when Russell himself made such remarks he almost invariably did so with ontological implications. This point, which I did not invent or discover but took from Moore and Stebbing, is abundantly clear in Russell's writings. When he says, for example, that Socrates is a logical construction, he is not merely making the point that, unlike 'this', 'Socrates' is not a logically proper name; he is propounding the ontological view that persons, unlike sense data, are not ultimate particulars. For Russell the doctrine of incomplete symbols fulfilled the role of an ontological *via media* between, on the one hand, admitting inferred entities and the paradoxical outright denial of the existence of numbers, classes, persons, physical objects, etc. on the other. Would Mr. Perkins wish to rephrase Russell's version of Occam's

Razor to read: do not refer to inferred entities by logically proper names, but indirectly by means of definite descriptions?

*Corpus Christi College, Oxford*

## HARMAN'S NONESSENTIAL PROPERTY

By GEORGE F. SCHUMM and JOHN G. STEVENSON

GILBERT Harman argues ('A Nonessential Property', *Journal of Philosophy*, vol. 67 (1970), pp. 183–185) that the claim that numbers have essential properties is incompatible with the idea that number theory can be reduced to set theory in a variety of ways:

A particular set will be a composite number given certain reductions but not others. Apart from one or another of these reductions, we cannot say that a particular set is or is not a composite number. If *de re* necessity is in question, no set is necessarily a composite number. Being a composite number is not an essential property of any set. Therefore, if numbers can be identified with sets and *de re* necessity is in question, no number is necessarily a composite number. Being a composite number is not an essential property of any number.

The argument rests upon an equivocation. If to say that numbers can be identified with sets is to say that numbers *really are* sets, as perhaps Frege thought, then there is at most one correct "reduction" of number theory to set theory—*viz.*, that which identifies each number with the set that it *is*—and not the multiplicity Harman needs in order to show that no set is essentially a composite number. On the other hand, he may mean to say only that number theory has a model in set theory. In that case, we have the requisite multiplicity of identifications; but it is now open to the essentialist to say simply that the question whether being a composite number is an essential property of any *set* has nothing to do with the question whether *numbers* have this property essentially. For it is only in the stronger sense of 'reduction', in which numbers are taken to be sets and are not merely modelled by them, that reduction of a number to a set requires that the number share all the properties of the set to which it is reduced.

The essentialist does indeed have his share of headaches, many of them well-deserved; but Harman's argument, at least, is no cause for pain.

*Ohio State University and Purdue University*

4 MAY 1973

## NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, R. G. Swinburne; J. H. Benson, Alec Fisher, Andrew Harrison, R. F. Holland, A. C. MacIntyre, Bernard Mayo, A. R. White, Peter Winch. This committee is responsible for appointing and advising the editor and for the general policy of the paper.

**SUBSCRIPTIONS.** The subscription to ANALYSIS, Vol. 33, is £1.50 for individuals in Great Britain, £1.75 for individuals abroad, and £2.00 for all institutions, post free. Each volume comprises six numbers, five of 32 pages and one of 48 pages, appearing within the academic year—in October, December, January, March, April and June. Orders should be sent to Basil Blackwell, 49 Broad Street, Oxford, or placed with any bookseller.

**CONTRIBUTIONS.** Articles submitted for publication should be addressed to Dr. C. J. F. Williams, 1 Fossefield Road, Midsomer Norton, Bath, BA3 4AS. Contributors are asked to note the following requirements.

Articles should normally not exceed 3,000 words in length. Occasionally, however, longer contributions can be accepted;

They must be typewritten in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgement is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope and international reply coupons of the requisite value, whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

The copyright of articles printed in ANALYSIS remains the property of the author, but contributions are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.

# ANALYSIS



Edited by  
C. J. F. WILLIAMS

---

## CONTENTS

Formal justice and rule-change	T. D. CAMPBELL
Two views of moral practices	CHRISTOPHER CHERRY
Some supposed advantages of materialism	O. H. GREEN
Aesthetic judgments and non-aesthetic conditions	GARY ISEMINGER
Proper names and <i>suppositio personalis</i>	BARRY MILLER
Davidson on saying that	WILLIAM G. LYCAN
Frege's way out: a footnote to a footnote 139	MICHAEL DUMMETT
How things exist: a difficulty	ROBIN ATTFIELD
Censorship, state secrets and participatory democracy	H. M. JONES

---

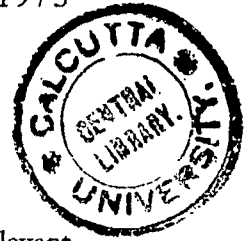
BASIL BLACKWELL · ALFRED STREET · OXFORD

---

*Forty pence*

## FORMAL JUSTICE AND RULE-CHANGE

By T. D. CAMPBELL



FORMAL justice requires the correct application of the relevant distributive rule to all cases which are alike within the terms of that rule; it is treating like cases alike where likeness is a function of the categories of person distinguished in the rule to be applied. From the moral point of view formal justice is significant where failure to apply the relevant rule affects the distribution of benefits and burdens as between those to whom the rule is correctly applied and those to whom it is not. The unfairness of formal injustice has nothing directly to do with the fairness or otherwise of the rule (i.e. material justice) but concerns the relative positions of those who are and those who are not treated in accordance with appropriate rule. It can arise therefore only within the framework of the rule of law, that is of a system of general rules which are taken to be authoritative for a given set of persons.

The nature of formal justice is normally illustrated by citing examples of the injustice of making individual exceptions to rules for reasons not provided for in an existing "higher" rule and by pointing to the judiciary's duty to be impartial in the administration of the law. Another aspect of formal justice is less commonly noted: if it is formally unjust not to apply the same rule to similar cases then it must follow that it is formally unjust to change a rule since this will involve treating persons who are in essentially the same category differently on the irrelevant grounds that the cases occurred at different times. Thus if rule 1 is applied before time  $t$  and rule 2 is applied after time  $t$ , in terms of rule 1 what happens after time  $t$  is formally unjust and in terms of rule 2 what happened before time  $t$  was formally unjust. And so, from whichever point in time we view it, rule-change involves formal injustice. In either case we would say that the injustice is done to the person or persons who are less well treated than they would have been under the rule which was not applied to their case.

That the difference in treatment is not the result of a single isolated act but follows from a change in the rule to be applied does not alter the fact that there is formal injustice as between group A, whose cases occurred before  $t$  and which were therefore treated under rule 1, and group B, whose cases arose after  $t$  and which were therefore treated under rule 2. Either group could claim that essentially similar cases (whether similarity is defined in terms of rule 1 or of rule 2) were being treated differently. In this important respect to change a rule is equivalent to making exceptions to it. For instance if group B suffers as a

result of the change in that they would have been better off under rule 1 than under rule 2, then whatever the material justice of the change, individuals within group B would have justifiable grounds for complaining that it is unfair that they should be relatively worse off than members of group A merely because their cases occurred at a different point in time. The same rule has not been applied to relevantly similar cases and this is formally unjust. Therefore when a change in a rule affects the distribution of benefits and burdens within a group living under the same system of rules then there is always at least one good moral reason against making such a change. This is the thesis which I wish to discuss.

# I

It might be said that there must be at least one important type of exception to this thesis, namely those cases where the existing rule explicitly contains a time clause, e.g. if rule 1 states that only group A, being persons in category C before  $t$ , come within its scope, then group B, being persons in category C after  $t$ , are specifically excluded from the rule and it is not therefore formally unjust to treat them in accordance with rule 2.

But there are two importantly different ways in which a rule may make explicit reference to groups of persons according to their position in time. The time referred to may be general or particular, that is, it may be a point in a repeating time scale such as 12 noon, 1st January or the end of a decade, or it may be a specific moment in time such as 1st July, 1972. Now it is a recognised feature of the rule of law and therefore of a system of rules compatible with formal justice that the rules in question should not make reference to particulars but should refer only to categories of persons, situations and treatments. Otherwise, from the point of view of the rule of law, they are not rules of the appropriate sort laying down different types of treatment for different types of persons. So also with respect to time references, where a rule contains a reference to what I have called a general time (e.g. no swearing on Sundays) this is perfectly compatible with the rule of law and with formal justice; but it has nothing to do with changing the rules. Where a particular time is mentioned in a rule as, for instance, where a time limit of a specific date is incorporated within it so that the rule is restricted to a named historical period, then formal injustice arises because of the lack of generality in the rule. From the point of view of the rule of law, rules containing a specific time reference are not genuinely general rules, and to apply such defective rules is tantamount to formal injustice since the distinction between group A and group B is arbitrary as opposed to rule-governed. I conclude therefore that we cannot formulate rules in such a manner as to make them compatible with the

rule of law and at the same time avoid the formal injustice which is consequent upon changing them.

## II

The thesis may be challenged by attempting to reduce it to absurdity. If, it could be argued, it is always unjust to change rules, then, this being an argument against all institutional changes, it must always be wrong to alter the status quo; and, since it is often indisputably desirable to change rules, an argument against all changes in rules can have no force.

But the thesis is not that all changes in rules are morally indefensible but that, whatever the moral reasons in favour of changing rules, such as considerations of material justice for instance, there is always at least one moral reason against such a change, namely that with respect to groups A and B it is formally unjust. It is not claimed that this is an overriding reason for not altering the status quo, but only that it is a *prima facie* one. For, even if justice is the over-riding moral requirement for a social rule or institution, so that to condemn a change as unjust is to put forward a decisive moral reason against it, nevertheless formal injustice is only one aspect—and not necessarily the most important aspect—of justice.

Further it should be noted that the thesis does not state that it is formally unjust to introduce a new rule to meet a new situation which has arisen in the course of social or economic development. In crude terms, if the new situation S did not occur before *t* then the fact that there was no such rule before *t* does not matter since, as there were no cases of S before *t*, this means that group A had no members and therefore formal injustice is avoided. In practice, of course, situations usually change gradually while rules can only be changed in discrete steps, so that formal injustice will tend to arise as between group B and those members of group A whose circumstances were affected by the gradual emergence of S before *t*. But in theory if new rules were introduced contemporaneously with the emergence of S then the change would not lead to formal injustice.

## III

Another line of attack on the thesis is to argue that all rules have an implicit qualifying clause saying that they are subject to change, because, it could be said, it is a generally understood fact, at least in a modern industrial society where social change is taken for granted, that no rules are to be thought of as equivalent to the laws of the Medes and Persians. If this is so, it is argued, then no person can complain when rules are changed because he is not entitled to expect that no such changes would be made.

The quick answer to such an argument is to point out that it is tantamount to saying that it is not unjust for one person to treat another

unfairly provided that the latter has been given reason to believe that this is likely to happen. But the declared intention to act unjustly could only marginally alter the moral character of an unjust act when it is done. So while it may be necessary to make frequent alterations in rules this should be recognised to be a tragic fact rather than a morally indifferent one.

This reply is, perhaps, too sweeping. We must distinguish between rules on the basis of which people make commitments and forward plans based on the expectation that the rules in question will continue, e.g. as when I take out a mortgage counting on the continuance of the law giving tax relief with respect to interest payments on house loans, and rules which typically do not govern people's forward planning in this way, e.g. the law providing for the punishment of murderers. I suppose that there are no rules which it is not possible to conceive of as inducing persons to enter into commitments in the expectation that they will continue, but we can distinguish between the injustice of altering a rule because the consequent legitimate expectations of people affected by the rule will be disappointed (which I will call *expectational formal injustice*) and the injustice of altering a rule which arises simply from the fact that this involves treating differently persons in the same essential category (which I will call *pure formal injustice*). The former could be avoided if it were explicitly stated or tacitly understood that persons affected by a rule should not count on its not being changed. The latter is not avoidable in this way.

While *expectational formal injustice* is clearly immoral in so far as it is equivalent to disadvantaging someone by breaking a promise on the basis of which they have entered into certain commitments, it might be argued that there is nothing essentially immoral in *pure formal injustice*, thus reducing the thesis to moral triviality. But this is not so. The immorality of changing the rules cannot be reduced to causing hardships by disappointing legitimate expectations (*expectational formal injustice*) or to the injustice of one or other of the rules (*material injustice*), for it concerns the irreducible unfairness of disadvantaging one group of people relative to another group who are in essentially the same category. Consistency in the application of rules distributing desired goods to human beings is no mere intellectual virtue, for it is an essential part of what is involved in treating them fairly. We sympathise with those who are disadvantaged by a failure to apply established rules and a similar sense of unfairness is evident when rules are changed in such a way that one group of persons gets an advantage or disadvantage that was denied to or avoided by equivalent persons before or after a change in the rules; consider for instance our feelings about those persons who are too old to qualify for an old age pension because state insurance schemes were introduced too late for them, or more generally the resentful attitude of



older persons when younger ones are given more advantages than they themselves enjoyed. This resentment may be diminished by the thought that the rules have become more humane or materially more just, but a sense of unfairness remains which cannot be dismissed as mere pique or envy. This is independent of the justice of the rules themselves, for even if we could know that the old rule was more just than the new one so that group A were treated justly in the material sense, there remains an unfairness as between those who were treated justly (materially) but more harshly and those who were treated unjustly (materially) but more leniently.

#### IV

But can the resentment felt by those who are adversely affected by pure formal injustice really be defended in cases where the rules are being made materially more just? Surely we could never use the argument that persons should be treated in the future in accordance with materially unjust rules because essentially similar persons have been so treated in the past? We can't be required to be materially unjust to group B in order to be formally just to group A.

Without coming to a view on whether or not material justice always overrides formal justice, we can note two points. First, if formal injustice is done for the sake of material justice, then it is a case of perpetrating certain injustice for the sake of uncertain justice, for while we can be sure that to change a rule is to alter the balance of benefits and burdens as between groups A and B, which is formally unjust, we cannot be sure that the new rule is materially more just. Now the difficulty of knowing what is materially just can hardly be used as a veto on all change but it is a point to be considered when there is disagreement about the material justice of the change in question. If there is unanimity on the material justice of the change then persons who are adversely affected by the change may be expected to acquiesce in it for the sake of the increase in material justice. But they could not be expected to do so where they thought the change did not serve the cause of material justice. Therefore where there is dispute about whether or not a change in the rules is a move towards or away from material justice then consideration of the formal injustice which would result from the change might tip the moral balance in favour of the status quo. For this reason we may say that, where the rule of law is established, formal justice always favours the established distributive order.

Secondly, even if it were accepted that material justice should always override formal justice in that the latter can never justify a failure to change the rules in order to render them more just, this would not necessarily make consideration of the formal injustice of rule-change otiose. It would provide one reason against the retroactive application

of a new rule when this would be to the disadvantage of group A, and one reason in favour of retroactive application of a new rule when it would favour the group treated under the old rule. Where this is not possible then formal justice suggests the need for compensation being paid to those whose relative position is adversely affected by the change in the rules. Or again the formal injustice of rule-change provides grounds for implementing such changes in a gradual manner so as to minimise the contrast in the treatment accorded to groups A and B. These considerations apply more strongly in the case of expectational formal justice but they are also relevant to pure formal justice.

I conclude therefore that, other things being equal, it is always *prima facie* unjust to change rules in such a way as to alter the relative distribution of advantages and disadvantages as between persons living under the same system of rules. Moreover this is no mere technical point but is a consideration which ought to be borne in mind either as an argument against change as such, or in relation to the way in which changes in rules are brought into effect.<sup>1</sup>

<sup>1</sup> I wish to thank Mr. A. J. M. McKay for his comments on an earlier draft of this paper.

*University of Glasgow*

## TWO VIEWS OF MORAL PRACTICES

By CHRISTOPHER CHERRY

### I

THE terms 'game', 'practice' and 'institution' feature prominently but untidily in a good deal of current moral philosophy. They tend to be put to much the same use; but it is not this unhelpful indiscriminateness that I want to discuss, and so I shall not watch my words too closely.

The concept of an institution, or practice, is cast in a leading role in one sort of attempt to establish a familiar thesis about the relationship between certain statements of fact and certain statements of value: the thesis that the former entail the latter. But what role, exactly? Writers appear to have proposed, without properly distinguishing, two ways of viewing and hence of characterising this role. Which of the two happens to be adopted, or at any rate emphasised, will depend upon what category of facts is taken as epistemological base-line. What base-line

and so which characterisation *ought* to be adopted is not, I shall suggest later, a matter of indifference. But my immediate object is first to describe, and then to indicate the consequences of confusing, the two viewpoints I have mentioned.

The notion of an institution, or practice, may be viewed (A) as a bridge the availability of which makes possible a deductive passage from statements of fact to statements of value—as a device which enables us to perform a logical operation with an impeccability which would otherwise be lacking. On the other hand, it can be viewed (B) as a tool for *elucidating* rather than *securing* an entailment relationship between such statements—as, so to speak, a marginal commentary upon an entailment whose existence has been denied or whose nature misrepresented.

The difference between (A) and (B) is well illustrated by an all too familiar example which fortunately can be given in an abridged form.<sup>1</sup> Consider the following statements:

- (1) Jones uttered the words 'I hereby promise to pay you, Smith, five dollars'.
- (2) Jones promised to pay Smith five dollars.
- (3) Jones placed himself under (incurred) an obligation to pay Smith five dollars.

Now, to adopt view point (A) is to focus attention upon the relationship between (1) and (3). To adopt viewpoint (B) is, by contrast, to focus attention upon the relationship between (2) and (3). One who stresses the *enabling* aspects of institutional concepts is attending to the relationship between brute facts and values, while one who stresses the *elucidatory* aspects of institutional concepts is attending to the relationship between institutional facts and values. And one who emphasises the first set of aspects to the exclusion of the second works with and from a base-line which differs from that of one who does not. It is not so much that he works with a model of the world as consisting of brute facts, and of knowledge as knowledge of brute facts, as Professor Searle would have it; it is rather that he recognises this, but no other, model as his jumping-off point.

The enabling device stressed by the first man is, in the present case, the institution, or practice, of promising. He sees the matter in the following terms:

- (4) In the context of the institution (practice) of promising, to utter the words 'I hereby promise to do X' is to incur an obligation to do X.

<sup>1</sup> Outraged purists should bear in mind throughout that this is no more than a skeleton. Reference to Professor J. R. Searle's famous article 'How to Derive "Ought" from "Is"' will rapidly provide flesh for the bones. See also Searle's book *Speech Acts*, in particular Chapter 8.

By contrast, the second, who stresses the elucidatory aspect, sees the matter thus:

- (5) By virtue of the fact that promising is an institutional concept, to promise to do X is to incur an obligation to do X.

It will be said that there is nothing *wrong* with either (4) or (5): different preoccupations engender different, but not mutually incompatible, formulations. But now imagine a third man who, failing to discern, or else acknowledge, any difference in preoccupation, borrows elements from both formulations. From (4) he borrows the enabling feature appropriate only to the first way of viewing things, namely the reference to context; and from (5) he borrows the reference to an institutional concept, a feature appropriate only to the second way of viewing things. Such promiscuous borrowing issues in the hybrid formula:

- (6) In the context of the institution (practice) P, to promise to do X is to incur an obligation to do X.

Now, what can P be? Obviously it cannot be promising as in (4), but some institution, or practice, within which promising can be executed, as promising is an institution, or practice, within which given utterances can be made. But if this is so, (6) at least suggests an analysis such as the following:

- (7) To promise to do X, in the context of institution (practice) P, as opposed to that of institution (practice) Q or to that of institution (practice) R, etc., etc., is to incur an obligation to do X.

(7) has rather curious implications. For one thing, it prompts such questions as: 'What, then, is promising like with respect to obligation outside the context of P? How does it work in the context of Q, for instance?' For another, it implies that promising is something which can be bodily lifted from a given institutional environment and scrutinized independently of any obligation-incurring feature. Whether or not promising possesses such a feature depends upon where it—the *same* it—happens to be planted. To move from a horticultural to a more grandiose spatial metaphor, (7) invites the wrong sort of Newtonian analysis of institutions, or practices. From the viewpoint of brute facts, institutions, or practices, do indeed look Newtonian: they are boxes in which brute facts may, and again may not, be gathered. But when we focus attention upon the obligations created by promise-making we are not adopting that point of view. Promising is not brute relative to incurring obligations as uttering the words 'I promise' is brute relative to promising. So if the logical space of promising is Newtonian with respect to certain utterances, there is nothing the logical space of which

is Newtonian with respect to the obligating features of promising. One who takes viewpoint (A) has Newtonian vision, and one who takes viewpoint (B) Leibnizian. Given their respective base-lines, their vision is sound. But one who attempts to combine both points of view achieves no clear vision at all.

## II

I want now to say something more about the two points of view I have distinguished; and I shall do so in connection with some remarks made by D. Z. Phillips and H. O. Mounce. In a book called *Moral Practices* they write:

Many philosophers, notably in this country Philippa Foot, who have maintained that a moral conclusion can be derived from a factual statement, have gone on to make far larger claims. They also want to maintain that a factual statement must entail the same conclusion for every moral agent, irrespective of the moral practices to which he belongs . . . for everyone who is a moral agent certain facts will entail that some things are right or wrong, but from this it by no means follows that the same facts will entail that the same things are right or wrong for everyone who is a moral agent. . . . To maintain that in a moral practice certain facts will entail certain moral conclusions does not preclude the possibility of there being different moral practices within which the same facts entail different conclusions. . . . The view that a given statement must entail the same conclusion for all moral agents is plausible only if we imagine a society that has a single moral practice or a set of moral practices that everyone shares. (Chapter 2, pp. 14-15).

Phillips and Mounce illustrate their thesis with promising. We are invited to imagine two societies: society A which has a 'single moral practice of promise keeping', and society B within which 'the concept of a promise does not have the place . . . that it has within the first [sc. A] society'. For the members of society A, the judgment that a person ought to repay money follows from his having said that he will do so. This is not so for the members of society B:

The facts which have a moral significance for the members of the first society will clearly not have the same significance for the members of the second. Thus, for a member of society B, the fact that he has said he will repay money he has borrowed will be a fact without moral import. . . . For such a person, who does not belong to the practice of promise keeping, saying that one will repay money has no relation to the judgment that one ought to repay it . . . there clearly is a sense in which, for a member of the second society, there exists a gap between his saying that he will repay the money and the judgment that the members of the first society would deduce from this. The view that a given fact must entail the same conclusions for all moral agents can be seen . . . to be a mistaken one. (*Op. cit.*, pp. 15-16.)

Phillips and Mounce conclude that since 'the concept of a promise does

not have the place within the second society that it has within the first, one would expect a member of the second society to consider it an arbitrary matter that the breaking of a promise is condemned . . . while he may understand what in a factual sense is involved in making a promise, the evaluative element in it will appear to him irrelevant.<sup>1</sup>

While it seems that Phillips and Mounce *intend* to take brute facts as their base-line and so to adopt the enabling view of practices, they sometimes switch to taking the institutional fact of promising as their base-line, with the result that the two points of view (A) and (B) are run together and obscurity supervenes. This they do because they take little trouble to distinguish the readily intelligible but uninteresting suggestion that society B has no practice of promising from the question-begging but more provocative suggestions that society B has a different practice of promising, or that it has a different attitude towards the practice of promising, from Society A. This carelessness leads to talk of factual and evaluative elements in promise-making, and thence to the necessity (and sufficiency) of some moral practice or other for a deductive passage from *promising* to obligation-incurring. This is as if one were to view soccer as *enabling* one to pass from 'offside' to 'free kick', or cricket as *enabling* one to pass from 'clean bowled' to 'out'. Offside and clean bowled are not, however, waiting to be made honest women of. There is a variety of joke which exploits the anachronistic and goes something like this: In 2000 B.C. one man is going through the motions of playing chess. His friend comes along and exclaims: 'Hey, you can't do that: chess hasn't been invented yet!' This could not be a joke, funny or otherwise, if practices—whether promising or soccer or cricket or chess—were viewable in the above way.

Let us now leave this muddle and take, as Phillips and Mounce must, brute facts as our base-line. For to make good their conclusion that it is mistaken to suppose either that a given fact must entail the same conclusion for all moral agents or that facts never entail moral conclusions, Phillips and Mounce must show that it is from the fact that N said he would repay money owed rather than from the fact that N *promised* to repay money owed that it follows in society A, but not in society B, that N has incurred an obligation to repay money owed; and furthermore that such a state of affairs is to be explained in terms of moral practices exploiting identical matters of fact in different ways. Promising will be one such practice although not itself similarly exploitable. The trouble is that at this point an issue becomes important in a way in which it was not previously: the issue of whether or not we may after all exercise discretion in our choice of epistemological base-line. Until now our principal concern has been with a clear recognition and consistent application of whichever is the preferred base-line. What concerns us

<sup>1</sup> *Op. cit.*, pp. 16–17.

now is whether one base-line rather than another *ought* to be preferred; for the acceptability of Phillips's and Mounce's real thesis depends upon the answer we give to this question. To the extent that we are free (although not obliged) to adopt a brute fact base-line we are free (although not obliged) to speak of one and the same fact, namely N's statement that he will repay money owed, entailing in situation (A) but not in situation (B) the statement that N has put himself under an obligation to repay money owed. But if we enjoy no such freedom we cannot speak in the sense demanded by the thesis of the *same* fact. We can of course make a sort of identity claim, for we can speak of the same utterance and perhaps of the same words (though not of the same *form* of words). However, this claim is too anaemic to support the weight of the thesis.

Let us make the example clearer by supposing that a member of (A) and a member of (B) both utter the words 'I will repay the money I owe you'. One who denies the requisite freedom will argue that, for all purposes presently relevant, there are two quite different facts: the fact that a man promised to repay money, and the fact that a man said that he would repay money. For whether or not the states of affairs are to be deemed the same depends upon the respective roles of the utterances in question; and this means that a determination of identity waits upon, and cannot be settled in advance of, an exploration of how each engages with questions of obligation. And they engage very differently. *Before* and not after we talk about sameness, we must be prepared to talk about moral import.

If the terms 'correct' and 'incorrect' are allowed here, then I think the arguments and the conclusion of the person who refuses this epistemological licence are correct and those of Phillips and Mounce incorrect. If they are disallowed, I am content to settle for 'profitable' and 'unprofitable'. There are more and less profitable ways of regarding conceptual machinery; and the machinery of moral practices or institutions is more profitably regarded as mining moral facts than as striking a bedrock of brute facts which, although always available, is bound to lie beneath what is really sought.

*University of Kent at Canterbury*

## SOME SUPPOSED ADVANTAGES OF MATERIALISM

By O. H. GREEN

THE primary concern of philosophers who defend materialism has been to show, as U.T. Place put it, 'that the thesis that consciousness is a process in the brain cannot be dismissed on logical grounds' ('Is Consciousness a Brain Process?' reprinted in V. C. Chappell, ed., *The Philosophy of Mind*, p. 102). Over and above this, they have largely been content to point out certain advantages which materialism appears to offer. The empirical basis which the theory requires has received little attention from its philosophical proponents; they leave such experimental confirmation as can be given the theory to physiologists. I wish to argue that a consideration of the empirical basis of materialism shows that some advantages claimed for materialism are supposed advantages only.

### I

The advantages which I have in mind concern the resolution of the problem of other minds and the elimination of troublesome psycho-physical connections. Materialism has frequently been taken to provide a straightforward, scientific account of knowledge of the minds of others. The details of the account vary somewhat. Sometimes appeal is made directly to the claimed identity of mental states with brain states; in other cases a causal analysis of mental concepts is employed as well. I shall refer to the former approach as the Basic Position and to the latter as the Causal Position.

Philosophers who take the Basic Position reason that if mental states are identical with brain states, all that is necessary to determine that another person has a certain mental state is to discover that his brain is in the relevant state, which it is possible for physiology to do. It is this approach which is set out in the following passages:

[According to materialism] a person's sensations are processes in his brain, no more and no less, and the occurrence of brain processes can be inferred by certifiable methods—the regular methods of scientific, or more particularly of physiological, inquiry. (V. C. Chappell, ed., *op. cit.*, Introduction, p. 19.)

Suppose . . . that the identity theory of mind and body is true and known to be so. . . . Then the epistemological problem of other minds has an easy solution. There would be reliable methods of determining whether that object over there has a mind because the occurrence of brain states can be determined by scientific methods—the methods of physiological inquiry. (James F. Tomberlin, 'Plantinga's Puzzles about God and Other Minds', *The Philosophical Forum*, I (1969), 384–85.)



The argument of philosophers who adopt the Causal Position is that mental states are the causes of certain patterns of behaviour, and physiology reveals that the causes of the behaviour are brain states, in this way providing a way of solving the problem of other minds. In the following passages it is this approach which is suggested:

Our [causal] account of the nature of mental states makes the problem of the existence of other minds particularly easy to solve ..... We need only three premisses to infer the existence of a mind that this behaviour is an expression of. (i) The behaviour has some cause; (ii) the cause lies in the behaving person; (iii) the cause is an 'adequate' cause—it has a complexity which corresponds to the complexity of the behaviour. (D. M. Armstrong, *A Materialist Theory of Mind*, pp. 123, 125.)

The causes of behaviour which are supposed to be mental states are, in Armstrong's view, states of the brain, and it is the business of physiology to discover them.

Central-State Materialists construe the problem of other minds as a problem in establishing like causes for like behavioural effects..... By showing that in both of us the causes of behaviour lie in the brain, we show both that he has a mind, as we do, and that he has a mind like ours. (Keith Campbell, *Body and Mind*, pp. 119–120.)

The second advantage claimed for materialism is that it eliminates the bothersome psycho-physical connections which clutter the landscape presented by interactionism, epiphenomenalism, and other dualistic theories. This point is given special emphasis by J. J. C. Smart, who argues for materialism in this way:

If it is agreed that there are no cogent philosophical arguments which force us into accepting dualism, and if the brain process theory and dualism are equally consistent with the facts, then the principles of parsimony and simplicity seem to me to decide overwhelmingly in favour of the brain process theory. [For] dualism involves a large number of irreducible psychophysical laws . . . of a queer sort. ('Sensations and Brain Processes', reprinted in Chappell, ed., *op. cit.*, p. 172.)

## II

The problem of other minds is, roughly, the problem of how one can know that others have thoughts and feelings. It is tempting to believe that if thoughts and feelings are identical with brain states, knowledge of other minds will be possible in just the way we know about other brains—by physiological examination. This is, of course, the Basic Position. It must be remembered, however, that we can take thoughts and feelings to be identical with brain states only if we know (among other things) that particular thoughts and feelings occur when, and only when, certain brain states do. The question thus arises as to how this knowledge is possible. In order to establish the necessary co-occurrence, it must be determined that thoughts and feelings occur at just the times

when the relevant brain states occur. In one's own case this could be done straightforwardly; it would simply involve doing a bit of introspection and some physiology. The times when one had a thought or feeling would be noted and matched with the times when brain states occurred. This technique could not be employed in the case of other people, however, and the introspective physiologist would be unpardonably hasty in generalising from his own case. In order to establish that there is a general correlation of mental states with brain states, it is necessary to determine that other people have thoughts and feelings independently of any consideration of their brain states. For, since the identity of brain states with mental states depends on their correlation, one can hardly rely on their identity in order to determine that they are correlated. The appeal to the identity of brain states with mental states fails to provide a solution to the problem of other minds because materialism can be known to be true only if an independent solution already exists.

At this point the difference between the Basic Position and the Causal Position becomes important. The Causal Position appears not to depend on the establishment of a correlation between mental states and brain states, but rather proceeds directly from a consideration of behaviour. The line of reasoning is as follows. An itch, which we may take as an example, is the cause of a certain kind of behaviour—scratching. From physiological investigations we find that a brain state causes this behaviour. We conclude that, as a matter of fact, an itch is a brain state.

The crucial question here is whether we can reasonably suppose that the cause of the behaviour is (simply) a brain state. It is a remarkable fact that, in ignorance of our brain states, we are able to ascribe itches to ourselves and to others and to explain our behaviour as being caused by such feelings. In doing these things we do not conceive of itches as bare causes. If we stimulate a man's brain so as to make him scratch, although he doesn't feel anything, we have induced in him a cause of scratching but not an itch. And when we say that a man scratched because he had an itch (and not because his brain had been tampered with, etc.), we are not just saying that whatever causes scratching caused him to scratch, which is no explanation. This is not to say that the concept of an itch is not (at least in part) to be explained in terms of the causation of scratching, and it is not to deny that brain states are involved in the causation of such behaviour. But it does mean that there is a good case for thinking that, even though brain states may be at least partial causes of the behaviour, certain feelings are also causally involved. A purely physiological consideration of the causes of behaviour will not give us knowledge of other persons' feelings.

To this conclusion the defender of the Causal Position will want to

object. He might argue, for one thing, that the occurrence of an appropriate brain state is sufficient to cause scratching, so that there is no reason to bring a supposed feeling into the causal explanation of the behaviour. But the only way to show this is to suppose the behaviour to result from the occurrence of the brain state in the absence of any feeling, and this turn of events would empirically falsify the materialist's thesis that feelings and brain states are identical.

The proponent of the Causal Position might also object that the feeling of an itch is identical with a brain state, so that a physiological study of the causes of the relevant behaviour is sufficient to obtain knowledge of other people's feelings. But what is the basis for this identity claim? It cannot be that the brain state alone accounts for the behaviour typical of having an itch, for that is just what is in doubt. There remains the possibility of basing the claim on the fact that the feelings occur when, and only when, the brain states do. But in that case we are back to the Basic Position, which presupposes an independent solution to the problem of our knowledge of the feelings of others.

### III

If mental states and brain states turned out to be identical, certain kinds of psycho-physical connections would be eliminated. There would hardly be any question of a causal connection between them, for example. But this does not mean that we would be in a position to eliminate all psycho-physical links. In order to see this we must consider again the empirical basis which materialism requires.

It is important to note that the identity of brain states and mental states cannot be as simply established as can many other cases of contingent identity. To show that Mr. Nixon is identical with the President of the United States (in 1972), it must be determined (among other things) that public appearances of the President coincide with those of Mr. Nixon. This can be ascertained by observations of a single sort—by seeing the man in question. The case is different when we set about showing that a thought or feeling is identical with a brain state. The occurrence of a brain state is determined by the use of instruments of appropriate kinds, but the occurrence of a thought or feeling is accessible in one's own case to introspection and in the case of others through observation of their behaviour, etc. No one observation, not even observations of the same kind, will be sufficient to show that mental states are correlated with brain states. The correlation will involve matching one set of observation-conditions, those relevant to brain states, with another, those appropriate to mental states. Thus, a link between physiological observation-conditions and psychological observation-conditions is required if materialism is to be shown to be true, even if other psycho-physical connections can be eliminated. The

world-view presented by materialism is, in this respect, not quite so streamlined as some of its advocates have supposed.

#### IV

The materialist theory that mental states are identical with brain states can be true only if the following principle, which we may call the Correlation Principle, is true: For any mental state *M*, there is a brain state *B* which obtains when, and only when, *M* does. Moreover, the truth of the materialist theory requires that the Correlation Principle be established empirically; for the identity of mental states with brain states is held to be, not a matter of *a priori* necessity, but a contingent matter of fact. The argument of this paper has been that in thinking that materialism offers the advantages of a simple solution to the problem of other minds and a world-view without the encumbrance of psycho-physical connections, philosophers overlook the dependence of the theory on the empirical confirmation of the Correlation Principle.

Materialists dislike talk of a correlation of states of consciousness with brain states, for, as Smart suggests, 'to say that they are *correlated* is to say that they are something "over and above". You cannot correlate something with itself.' ('Sensations and Brain Processes', reprinted in V. C. Chappell, ed., *op. cit.*, p. 161.) Thus it may be objected that it is question-begging to saddle the materialist with the Correlation Principle. I do not think that this is the case, however. The identity of mental states and brain states, if it obtains, cannot be discovered *a priori*. Establishing it involves showing that certain conditions are met, including the co-occurrence of the putatively identical states. This, in turn, requires that both physiological and psychological observations be made. Only this is required for the argument. The question of whether mental states and brain states are identical is left open. It is certain advantages of the identification, supposing that it is possible, which are called into question.

Remarks made by Smart and Armstrong suggest that materialism may afford an insight about other minds and a simpler world-view in a way which the argument seems to overlook. It is reasonable to suppose, they suggest, that examination of a person's brain might come to constitute a criterion for his being in a mental state, and, indeed, might become the standard way of determining his state of mind. If this is so, then no reliance on psychological considerations or correlations of observation-conditions is required. Here it must be stressed, however, that before this hegemony of physiology could be established, the necessary psycho-physical correlations would have to be discovered independently. It is also worth noting that if brain states were criteria for mental states, this would mean that the concept of a mental state would be explained, in part at least, by reference to brain states. The

identity between mental states and brain states would in that case no longer be contingent.

*Tulane University*

## AESTHETIC JUDGMENTS AND NON-AESTHETIC CONDITIONS

By GARY ISEMINGER

IN his widely-discussed article 'Aesthetic Concepts' (*The Philosophical Review*, Vol. 68, 1959) Frank Sibley makes two claims which, given an initially plausible assumption, appear to be contradictory. He repeats these claims in later articles ('Aesthetic Concepts: A Rejoinder', *The Philosophical Review*, Vol. 72, 1963, p. 82, and 'Aesthetic and Non-aesthetic', *The Philosophical Review*, Vol. 74, 1965, p. 152); so far as I can tell, they are not significantly altered in these later formulations. (In what follows, numbers in parentheses refer to pagination in W. E. Kennick, *Art and Philosophy*, New York, 1964, which contains a revised version of the original article.) The two claims are:

- (1) There are no non-aesthetic features which serve in *any* circumstances as logically *sufficient conditions* for applying aesthetic terms (353).
- (2) [Aesthetic terms] may be governed *negatively* by conditions (355).

In order to see the problem which these claims create, we must first understand them and the terms used to make them. 'When a word or expression is such that taste or perceptiveness is required in order to apply it, I shall call it an aesthetic term or expression. . .' (351). Examples of aesthetic terms cited by Sibley are: '*unified, balanced, integrated, lifeless, serene, sombre, dynamic, powerful, vivid, delicate, moving, trite, sentimental, tragic*' (351). I think we may paraphrase (1) in the following fashion, where 'N' ranges over non-aesthetic features and 'A' over the aesthetic features which correspond to aesthetic terms:

- (1a) For all N and for all A, it is not the case that anything's having the feature N entails its having the feature A.

The concept of being 'governed negatively' by conditions is illustrated by the following examples, among others: 'It may be impossible that a thing should be garish if all its colours are pale pastels, or flamboyant if all its lines are straight' (355). For a term to be governed negatively by

conditions, then, is for it to be such that there are conditions whose fulfilment entails that it is not the case that the term applies. I think then that we may paraphrase (2) thus:

(2a) There is an  $N$  and there is an  $A$  such that anything's having the feature  $N$  entails its *not* having the feature  $A$ .

In saying that (2) can be paraphrased as (2a), I take the word 'may' in the preceding quotation to express an unwillingness on Sibley's part to commit himself to just this example rather than any modal notion involved in the claim. The word 'may' continues to appear, however, in the most recent of the articles cited: 'No nonaesthetic conditions or descriptions logically require the application, though some may require the rejection, of an aesthetic term' (*The Philosophical Review*, Vol. 74, 1965, p. 152). For this reason, it is perhaps worth mentioning possible modal interpretations of (2). There are at least two modal claims which Sibley might be making:

(2b) It's possible that there is an  $N$  and there is an  $A$  such that anything's having the feature  $N$  entails its *not* having the feature  $A$ .

(2c) There is an  $N$  and there is an  $A$  such that it's possible that anything's having the feature  $N$  entails its *not* having the feature  $A$ .

In the first place, however, neither (2b) nor (2c) seems to be of much interest for aesthetics, since neither would entitle us to conclude, given the presence of a non-aesthetic feature which stood in the specified relation to an aesthetic feature, that the aesthetic feature was absent. In the second place, given plausible though by no means indisputable modal principles, the argument I shall offer showing that (2a) entails the denial of (1a) can be extended to show that each of (2b) and (2c) entails the denial of (1a) as well. (The modal principles in question are the 'strong reduction principle' and laws involving relations among modalities and quantifiers due to Ruth Barcan Marcus. See W. and M. Kneale, *The Development of Logic*, Oxford, 1962, pp. 551, 614.) I shall, therefore, take (2a) as the correct expression of Sibley's second claim.

To show that (2a) entails the denial of (1a), let us follow Sibley's example and let  $N$  be the feature of having all straight lines and  $A$  be the feature of being flamboyant, where something's having  $N$  entails its not having  $A$ . (Whether this entailment holds for *these* choices of  $N$  and  $A$  is irrelevant to my purposes, so long as, in accordance with (2a), there *are* choices for which it does hold. Any such choices will do.) We then have:

(a) Anything's having all straight lines entails its not being flamboyant.

Invoking obvious logical principles relating negations of propositions to complements of terms and sanctioning the substitution of logical equivalents for one another, we may infer

- (b) Anything's having all straight lines entails its being non-flamboyant.

If we invoke a further, no doubt less obvious, principle,

- (3) The complement of an aesthetic term is an aesthetic term,

then (b) clearly entails

- (c) There is an *N* and there is an *A* such that anything's having the feature *N* entails its having the feature *A*.

which entails the denial of (1a).

How might one attack (3)? One might deny that any negative terms are aesthetic terms. But, quite apart from the theoretical difficulties of dividing terms, considered by themselves, into negative and positive terms (which may be contrasted with the theoretical ease of determining whether or not two terms are related to one another as complements), at least one of Sibley's examples (*lifeless*) would seem to be a negative term if any sense can be made of the notion at all.

A more promising move would be to suggest that in cases where an aesthetic term *is* governed negatively by conditions, its complement is *not* an aesthetic term. Something like this would seem plausible in the present case, since we can take it that it does *not* require taste or perceptiveness in order to apply the term, 'having all straight lines', and it certainly does not require taste or perceptiveness to recognize that having all straight lines entails being non-flamboyant, if indeed it does. To apply the complement of 'flamboyant', then, requiring only the ability to recognize a non-aesthetic feature and to understand a non-formal entailment, does not require taste or perceptiveness; the complement of the aesthetic term 'flamboyant' is, hence, not itself an aesthetic term.

In order to drive this wedge between some aesthetic terms and their complements, we must, of course, be careful to say what Sibley does in fact say, that an aesthetic term is a term such that taste or perceptiveness is required in order to *apply* it and not that an aesthetic term is a term such that taste or perceptiveness is required in order to *apply or withhold* it. Only if we avoid this latter claim can (3) be denied and the argument to show that (2a) entails the denial of (1a) be shown to be unsound. If we do thus deny (3), however, we must qualify a central claim which Sibley makes,

- (4) It is a characteristic and essential feature of judgments which

employ an aesthetic term that they cannot be made by appealing, in the sense explained, to non-aesthetic conditions (362).

For to judge that an aesthetic term does *not* apply is surely to make a judgment employing an aesthetic term. Further, one example Sibley gives of a judgment of which 'it would be natural enough to say' that making it 'requires the exercise of taste', an example which he then uses to illustrate the notion of an aesthetic term, is the judgment 'that a picture *lacks* [my emphasis] balance' (351). But where an aesthetic term which is negatively governed by conditions (and 'balanced', in fact, seems to be a plausible candidate for that status) is judged *not* to apply, the judgment that it does not *can* be made 'by appealing, in the sense explained, to non-aesthetic conditions'. That is, it can be made by noticing, without employing taste or perceptiveness, that the non-aesthetic conditions obtain and by inferring, in virtue of the fact that the presence of these conditions *entails* the inapplicability of the aesthetic term, that the term does not apply.<sup>1</sup>

<sup>1</sup> I am grateful to my students in aesthetics at Carleton College, especially Ruth Mattern, for discussions in which some of the points in this paper emerged, and to my colleagues, P. C. Mason of Carleton College and William Mann of St. Olaf College for comments on an earlier version.

*Carleton College*



## PROPER NAMES AND *SUPPOSITIO PERSONALIS*

By BARRY MILLER

PROPER names may be used so as to be self-referring, as in '“Tom” is a three-lettered word' and '“Tom” is a noun'. They may also be used to refer to their bearer, as in 'Tom is running'. And the point I want to discuss is whether 'Tom' may not be used in a third way, one in which it refers neither to itself nor to its bearer. What I have in mind may be illustrated from the case of common names. As a parallel to '“Tom” is a three-lettered word' and '“Tom” is a noun' we have '“Man” is a three-lettered word' and '“Man” is a noun' respectively. As a parallel to 'Tom is running' we have 'A man is running'. But, in addition, we have 'Man is a species', in which 'man' clearly refers neither to itself nor to individual men. So the question I am raising is whether there is any use of 'Tom' that parallels this last use of 'man'.

In the middle ages that same question was framed in terms of whether proper names did or did not have *suppositio personalis*, with William of Sherwood saying they did and Peter of Spain saying they did not. I think that William was right. I will argue therefore that there is a use of 'Tom' that does parallel that of 'man' in 'Man is a species', and in which 'Tom' refers neither to itself nor to Tom. This, I suggest, is to be found in 'Tom is an individual'.

Let me forestall possible objections by stating that 'individual' is being taken here as a second-order term<sup>1</sup> (correlative to 'species'), and is therefore *not* being used in any of the following senses:

1. as a synonym for 'chap' or 'bloke', as in 'Who is that individual there?', or in 'He is a hopeless individual'.
2. as a synonym for 'individualist', as in 'He is a real individual'.
3. as contrasted with 'being the holder of an office or a member of a group', as in 'He spoke as an individual, not as Prime Minister (or as a member of Parliament)'.

When 'individual' in 'Tom is an individual' is restricted to functioning as a second-order term, we could allow 'Tom' to refer to Tom only by allowing that a second-order predicate could be predicated of a particular—and that is not possible.

That might seem to settle the matter, were it only not so counter-intuitive; for when I say 'Tom is an individual' I do regard myself as saying something, no matter how trivial, about Tom and not about

<sup>1</sup> In this article I use the expressions 'second-order term' and 'second-order predicate' in the mediaeval rather than the Fregean sense. For Frege a second-order predicate is a predicate of a (first-order) predicate: it is said of such a predicate *qua* predicate. In the mediaeval sense, however, a second-order term or predicate is a sign of a sign: it is applicable to a sign *qua* sign.

'Tom'. In addition, there seems to be a distinct difference between 'Tom is an individual' and 'Man is a species'. We can argue, 'Tom is an individual. This is Tom. Therefore this is an individual'. However, we cannot argue, 'Man is a species. This is a man. Therefore this is a species'. Whereas there are clearly four terms in the second syllogism, it is by no means clear that there are four in the first. On the contrary, 'Tom' seems to be used in just one way, the same in both premises. And hence, since 'Tom' undoubtedly refers to Tom in the second premise, it should do so in the first as well. But if we admit that, we seem to be once more open to the charge of applying a second-order predicate to a particular.

Admittedly, that charge can be avoided in the case of at least two interpretations of the argument, 'Tom is an individual. This is Tom. Therefore this is an individual'. In the first of these interpretations 'individual' is to be taken in sense (1), i.e. to mean 'chap' or 'bloke'. But this is scarcely a solution, for it takes '— is an individual' as a first-order predicate, whereas we are discussing its use as a second-order predicate.

The second interpretation has the merit of accommodating that point. It consists in rendering 'Tom is an individual' by '“Tom” is an individual-designator'. It has the additional merit of showing that, like the syllogism in terms of 'species', that in terms of 'individual' also has four terms. It would run,

'Tom' is an individual-designator  
This is Tom  
Therefore, this is an individual-designator.

But this too is no solution, the reason being that '“Tom” is an individual-designator' is not synonymous with 'Tom is an individual'. If they were synonymous they should remain so even when 'real' (as opposed to 'fictitious') is inserted in each before 'individual'. That is, (A) '“Tom” is a designator of real individuals' should be synonymous with (B) 'Tom is a real individual'.

In (B) the use of 'real' shows that 'Tom' is a genuine proper name and not a shorthand description, as are the "proper names" in fiction. But if used as a proper name, it must, as Geach has argued, have associated with it a criterion of identity such that Tom may be said to be 'the same such-and-such (e.g. man)'. Hence 'Tom is a real individual' could be rendered as,

The man designated by 'Tom' is a real individual,  
or  
( $\exists x$ ) ( $x$  is a man &  $x$  is designated by 'Tom').

The analysis of (A) turns out rather differently, however. Since 'Tom'

is not used, but merely mentioned, it remains an open question whether 'Tom' has or has not been applied to anything: in fact, there is no sound whatever that could be precluded from being called 'a designator of real individuals', even though it may never actually have been used in that way. Since (A) is consistent with either situation, it implies neither. But if 'Tom' has not been applied to anything, it is not required to be associated with any criterion of identity. Hence (A) can be rendered as

$(\exists x)(x \text{ is designated by "Tom"})$ ,

thus showing that '“Tom” is a designator of real individuals' is not synonymous with 'Tom is a real individual'. To substitute '“Tom” is an individual-designator' for 'Tom is an individual' is therefore no way out of the earlier impasse.

However, there may still be a way out, if we reflect on the use of 'man' in 'Man is a species'. This does not mean '(The word, noun, name) “man” is a species'. Equally, it should now be clear, it does not mean '“Man” is a species-designator'. And indeed, if it were telling us something purely linguistic or logical, it would be written '“Man” is a species'. But although it is not doing that, and despite the lack of quotation marks around the grammatical subject of 'Man is a species', it is not strictly about individuals, i.e. 'man' is not being used as a common name. It seems therefore that its interpretation must occupy some middle ground—neither purely linguistic or logical on the one hand nor purely ontological on the other. And that is why I suggest that it be translated as follows:

Every man is (in English) called 'man', and is so constituted that more than one man may without ambiguity be called 'man'.

On this rendering 'Man is a species' is seen to be partly a logical proposition and partly an ontological one. Alternatively, it might be regarded as a logical proposition with an ontological basis.

Let me say why I have included the phrases 'is so constituted that' and 'without ambiguity'. Without 'is so constituted that' the translation would read 'Every man is (in English) called “man”, and more than one may without ambiguity be called “man”'. As thus stated, it is open to the interpretation that the possibility of applying 'man' to more than one man has no ontological basis, i.e. is not grounded in the kind of thing that each man is. Inclusion of 'is so constituted that' makes it quite explicit that it does have an ontological basis: it is grounded in the kind of thing each man is.

If 'without ambiguity' were omitted, the way would be open to the use of different senses of the one term for an individual of the one kind. I know of no example involving 'man', but here is one involving 'tick': 'This tick draws blood and that tick draws blood', where 'tick' means

'parasitic insect' in the first instance but 'sound of a clock' in the second. Of course this is nonsense, but that is not the point. The point is that, although there is nothing about a tick that requires it to be called 'tick', once it has been so called, the use of 'tick' for other kinds of things gives rise to ambiguity. And it is only for unambiguous uses of 'tick' that it would be true to say 'Tick is a species'.

If all this is correct, we need no longer be puzzled by 'Tom is an individual', for, parallelling 'Man is a species', it should be rendered as

A man is (in English) called 'Tom' and is so constituted that only he may without ambiguity be called 'Tom'.

What might be questioned is the inclusion of 'man': why should we say 'A *man* is (in English) called "Tom" . . . ? The reason is that 'Tom' is *ex hypothesi* a genuine proper name and not simply a fictional one. As a fictional one it would be simply a disguised description, but as a genuine proper name it is not reducible to any description (even a unique one) nor to any predicate (even Quinean ones like 'tomizes'). Ultimately it can have been introduced only in an act of naming: such an act is a necessary condition of a term's being a genuine proper name.<sup>1</sup> But if one could know nothing about the *kind* of thing to which one claimed to have given a proper name, one could not be said to have named anything. Hence, as Geach has noted, there is a common noun corresponding to every proper name. In the case of 'Tom' I am taking that noun to be 'man', although it could equally have been 'cockatoo', if I had named my pet cockatoo 'Tom'.

In connection with my suggested rendering of 'Tom is an individual' three things should now be noted. The most obvious of course is that there is no question of a category mistake in 'Tom is an individual', even though '—is an individual' is being used as a second level predicate.

The second is that the ambiguity in 'Tom is an individual' cannot be removed by enclosing 'Tom' in quotes. That much, at least, is clear from the lack of synonymy between 'Tom is an individual' and ' "Tom" is an individual-designator'. The quotation mark device suffices in such cases as ' "Tom" is a name' or ' "Tom" is a three-lettered word'. But 'Tom' is no more an individual than 'man' is a species. The suggested translations show that whilst 'Tom is an individual' and 'Man is species' may be about the terms 'Tom' and 'man', they are not simply about those terms. And hence it would be an over-simplification to attribute the ambiguity to absence of quotation marks. To do so would be to leave out of account all reference to the ontological grounding of 'Tom' and 'man'.

<sup>1</sup> Subsequently to the act of naming, it can of course be introduced by means of some description. But that it is not reducible to such a description is evident from the fact that the description may later cease to be true without at all making it improper to use the name introduced by it.

To admit this is not, of course, to countenance the reification of species nor of individuality, any more than to allow the ontological grounding of 'spherical' in 'The moon is spherical' is to admit the reification of sphericity.

The third point is that it is now clear that 'Tom is an individual. This is Tom. Therefore this is an individual' is an invalid syllogism, and hence does parallel 'Man is a species. This is a man. Therefore this is a species'. The two premises would be,

A man is (in English) called 'Tom', and is so constituted that only he may without ambiguity be called 'Tom'. But this is Tom.<sup>1</sup>

From which it follows:

Therefore this is so constituted that in principle only he may without ambiguity be called 'Tom'.

But that conclusion is not 'This is an individual', for it would then be rendered,

A man is called 'this' and is so constituted that in principle only he may without ambiguity be called 'this'.

The puzzles generated by 'Tom is an individual' have been resolved by these last three points. This makes it possible to say that 'Tom' may indeed be used in such a way as to be neither self-referring nor referring to Tom. William of Sherwood had noted much the same when he remarked that it 'is not the fact that an individual is supposed that produces personal supposition, but the fact that a thing bearing the form signified by the name is supposed, and this can occur in a proper name when it signifies a substance together with a quality. Thus when I say "Socrates is running" it is with respect to his real being (*respicitur pro sua re*); when I say "Socrates is predicable of one only" it is with respect to the form signified by the name'.<sup>2</sup>

<sup>1</sup> 'This' is not a name, and hence is not functioning as a *logical* subject. Nor, even though one could not use 'Tom' as a proper name unless one knew the kind of thing Tom is (e.g. a man), is 'This is Tom' simply elliptical for 'This man is Tom'.

<sup>2</sup> N. Kretzmann (translator and editor), *William of Sherwood's Introduction to Logic*, Minneapolis, 1966, p. 110.

## DAVIDSON ON SAYING THAT

By WILLIAM G. LYCAN

IN his paper, 'On Saying That' (*Synthese*, 19 (1968-69), pp. 130-46), Donald Davidson offers an analysis of indirect quotation, with the intention of displaying the logical form of any sentence instantiating the schema, '*S* said that *p*'. He construes the 'that' in such a sentence as being a demonstrative which picks out the speaker's *own* utterance of the sentence which replaces '*p*'. Thus,

(1) Galileo said that the earth moves

is to be analyzed as

(2) The earth moves. Galileo said that.

(2), Davidson says, is a 'definitional abbreviation' (p. 142) of

(3) The earth moves. ( $\exists x$ ) (Galileo's utterance *x* and my last utterance make us samesayers).

'That' in (2) apparently abbreviates the more elaborate referring expression 'my last utterance' in (3).

Galileo and I are 'samesayers' just in case 'an utterance of mine matches an utterance of his in purport' (p. 141). I find this relation obscure, but it is not my purpose to criticize it here. What I do want to argue is that (3) has at least one important consequence that (1) evidently does not have, and that therefore (3) is not the correct analysis of (1).

Presumably, 'my last utterance' in (3), being an indexical which picks out the immediately previous utterance of 'The earth moves', is a singular term in referring position (there is no suggestion on anyone's part that it occurs within an opaque construction). If so, we may safely apply existential generalization to it, and thus eventually derive

(4) ( $\exists x$ ) ( $x$ =my last utterance)

from (3).

But if (3) is the correct analysis of (gives the logical form of) (1), then whatever is entailed by (3) is entailed by (1). (For Davidson, to assign a logical form to a natural sentence is to provide a paraphrase of that sentence, which paraphrase is a truth-conditional structure, in that it can be given a formal truth definition of Tarski's sort. An analysis of a sentence thus wears its truth-conditions, and consequently its entailment-relations, on its sleeve; and the capturing of all and only the intuitively correct felt implications is a primary goal of the search for the sentence's correct analysis. (See 'The Logical Form of Action Sentences', in Rescher (ed.), *The Logic of Decision and Action*, University of Pittsburgh

Press, 1967, pp. 81–95.)) So if (3) entails the existence of *my* (the speaker's) last utterance of 'the earth moves', so does (1). (1), then, according to Davidson's analysis, is an indexical sentence which is ontologically committed (whenever a speaker utters it) to a current utterance of the speaker's. E.g., if I accept and utter (1) in 1972, I commit myself by accepting (1) to the existence of a certain utterance dated 1972.

But this cannot be right. (1) taken by itself does not entail anything about any 1972 utterance; Galileo could have said anything he liked without there having been any utterances (or even any people) after 1900. I conclude that Davidson's analysis is to this extent inadequate, since it attempts to saddle (1) with a consequence that (1) evidently does not possess.

*The Ohio State University*

## FREGE'S WAY OUT: A FOOTNOTE TO A FOOTNOTE

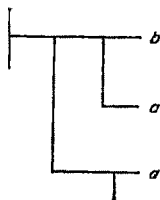
By MICHAEL DUMMETT

HAVING claimed that Frege's revised system was inconsistent, I went on to say that Frege may never have discovered this fact, but must have discovered that crucial theorems, such as that asserting the infinity of the series of natural numbers, could no longer be proved. Messrs. Linsky and Schumm (ANALYSIS 32.1, pp. 5–7) point out that, if the system was inconsistent, anything could be proved in it. I plead guilty to expressing myself carelessly. What Frege must surely have discovered is that the proof of the infinity of the number-series which he had previously given would no longer go through: it is further plausible that there is no formal proof of this theorem in the revised system which would appear intuitively convincing. In order to show that each natural number  $n$  has a successor, Frege proves by induction that, for each  $n$ , the number of numbers  $\leq n$  is a successor of  $n$ . It is easily seen that, in the revised system, the proof no longer goes through, since it cannot even be proved, in the same way as before, that  $0 \neq 1$ . Linsky and Schumm show that, from the revised axiom V', one can actually prove  $0=1$ . This, if Frege had realised it, would have shown him, not merely that his original proofs did not go through, but that the definition of 'number' would have to be revised. I was interested in answering the historical question why Frege never published the third volume of *Grundgesetze*, and gave up logical work after 1903. My answer is in no way impugned by the cavils of Linsky and Schumm.

Linsky and Schumm assert that Frege's revised system is consistent. Their ground is that Axiom V is the only non-logical axiom, and that the revised Axiom V' has a model in a one-element domain. This is to assume that Frege's 'logical' axioms I-IV are themselves satisfiable in such a domain. This is not the case, since Frege takes truth-values to be objects, and hence there must be at least two objects: Corollary IIIg of Axiom III reads (§50, Vol. I, p. 67):

$$\vdash (\neg a) = (\neg \neg a)$$

i.e.  $a \neq \neg a$ . Hence from the consequence  $a=b$  of the revised axiom, we obtain a contradiction. In virtue of Corollary Ia of Axiom I, which reads (§49, Vol. I, p. 65):



i.e.  $\neg a \rightarrow (a \rightarrow b)$ , the inconsistency involves, not merely the provability of a sentence and its negation, but the provability of every sentence. Thus it is entirely correct to assert, contrary to the opinion of Linsky and Schumm, that Frege's revised system is formally inconsistent.

*All Souls College, Oxford*



## HOW THINGS EXIST: A DIFFICULTY

By ROBIN ATTFIELD

IN his paper 'Aristotle on the Snares of Ontology' (*New Essays on Plato and Aristotle*, edited by Renford Bambrough, London, 1965) Professor G. E. L. Owen distinguishes three different existential uses of the verb 'to be', and defends an analysis which derives from Aristotle so long as it applies only to one of them, and to uses of sentences whose logical subjects are individuals. The first use occurs in 'Arrowby no longer is'; Owen decorates 'is' when thus used with a single asterisk. The second use is 'the use which is rendered by "il y a" or "es gibt" and represented in predicate logic by the formula " $(\exists x) Fx$ "' (ibid., pp. 84-5): this 'is' is distinguished by two asterisks. The third use is present in the assertion that time exists: it is no part of my present concern.

On the Aristotelian analysis discussed by Owen to be is always to be something or other. For Socrates, to be is to be a living man. A large number of different paraphrases will therefore be called for depending on what kind of thing is said to exist: 'exists' will be analysed as a predicate whose sense varies with the nature of the subject concerned. Owen wisely restricts the application of this analysis to 'exists\*', as opposed to 'exists\*\*', and also to singular propositions. Only thus does the analysis turn out to be defensible. Thus for the ice on a particular pond still to exist\* is for it still to be frozen water. But the defence which Owen conducts leaves doubts.

Let us look at the proposed analysis more closely. On this view the common feature of assertions of existence\* is the assertion of some referent or referents that it is or they are what it is to be of a certain sort. Assertions of existence\* observe, in fact, a common formula: the existence\* of a tree is to what it is to be a tree as the existence\* of a man is to what it is to be a man. For a man still to be\*, then, is for him still to be a man.

How then would Owen wish us to construe the assertion that a particular man is no longer in existence? For a man *not* still to be\* would be for him *not* still to be what it is to be a living man. But which individual is here referred to by the pronoun 'him'? Not, it would seem, a man; and no other sortal classification has been supplied. But if the term 'him' cannot be understood, the proposed analysis itself ceases to be comprehensible.

Owen would (see p. 79) here reply that for a man not still to be\* would be for some man who was at one time what it is to be a man no longer to be what it is to be a man. And this would elucidate the 'him' of the previous paragraph: the word 'him' goes proxy for '(the said) man who was at one time what it is to be a man'. We *can* refer to former

persons: Napoleon lost the Battle of Waterloo and is dead. 'Arrowby no longer is' will thus mean 'Arrowby was once a man, but is so no longer'.

The oddity of Owen's analysis is therefore yet to seek. It lies, I believe, in its purporting to translate what is not tautologous into a tautology. 'The ice is still in existence' will mean 'The ice is still frozen water'. But if to be ice is to be frozen water, then any ice is (still, now, or at any time) frozen water. The analysis, then, expresses a proposition which is necessarily true, whereas the analysandum in this case does not. It is not as if the ice might not have been frozen water. If ice is ice, it is frozen water, and where there is no frozen water there is no ice. Owen, at pp. 80, f., denies that there is any such problem over a particular patch of ice, as opposed to ice in general; but the distinction fails to rescue his analysis for the reason already indicated.

What was wrong with Owen's analysis of 'Socrates no longer is\*' as 'Socrates is no longer a man' was not that Socrates cannot now be referred to but rather the self-contradiction the analysis gives rise to in what was intended to be an analysis of a sentence which is not self-contradictory. For the name Socrates is only understandable aright if understood as the name of some man. Therefore the analysis involves the claim that a man is no longer a man. But there is nothing else that a man could be. A man can of course cease to be *alive*; but whenever even a former man is referred to, he is and must remain, so long as the reference is successful, a man. The name Socrates has reference even after Socrates drank hemlock; and the referent is a man, or nothing.

We are now in a position to return to the problem of the word 'him' in the proposed analysis, whereby for a particular man no longer to be\* is for *him* no longer to be a man. The problem does not turn on the impossibility of referring to individuals no longer in existence. It turns rather on making a successful reference to such an individual while at the same time predicating of it not being F, where its being F is a necessary condition of successful reference to it in the first place. We can indeed refer to Socrates, so long as we do not deny that he is a man: but if we do, no sense attaches any longer to our use of the pronoun 'him'.

My difficulty with Owen's analysis is thus that it represents certain contingent affirmative propositions as tautologies and certain contingent negative propositions as self-contradictions. But if asked to construct a better analysis I am myself in difficulty.

All I can say is this. For a man to be in existence is for him to live (though not to thrive). For inanimate objects to be in existence is for them to remain intact and, if they are functional, to retain at any rate residual powers of carrying out their function. A variety of analyses does accordingly seem called for of being\*. What relates these diverse analyses? Perhaps it is the common formula noted above as an implica-

tion of Owen's analysis: the existence\* of a tree is to what it is to be a tree as the existence\* of a man is to what it is to be a man and, come to that, as the existence\* of the Middle Ages is to what it is to be the Middle Ages.

But how is this formula to be cashed? What is the appropriate ratio in each case? I do not know. This is the nettle I have to leave for others to grasp.

*University College, Cardiff, and University of Ife, Nigeria.*

## CENSORSHIP, STATE SECRETS AND PARTICIPATORY DEMOCRACY

By H. M. JONES

PROFESSOR McMullen (in *ANALYSIS*, June 1972) draws attention to what he calls a 'paradox' about censorship and participatory democracy, though it is more a problem than a paradox. In any case, he provides the answer himself when he says: 'the organisation could decide that some or all of the offending material was so odious that they would ban it henceforth (either for all their number or for those presumed likely to be harmed further by it)'. This is the answer; it is no more problematic than a particular man's decision to refrain from actions in future that have caused him distress in the past.

The more important problem raised by Professor McMullen concerns state secrets. It is clear from a practical point of view that no such secrets could exist in an informed extreme participatory democracy (i.e.p.d.). Foreign powers find it easy enough to obtain our secret information now; with millions of people in the know they would have no difficulty at all. One might as well try to keep the English language a state secret. Democracy entails free discussion in free mass media: any foreign power could pick up our broadcasts; any visitor from abroad could read our newspapers.

But on examination the problem dissolves. The essence of a *state* secret, on our usage of this term, is just that it *is* known only to a very few, or at least (when there are leaks) that it ought not to have become known to the many. That is, 'state' in 'state secret' refers our attention to the small committee of pundits that actually runs *our* society; we would have state secrets if we were the Earth's only nation. That the notion of state secrecy has no necessary connexion with information

that might be useful to an enemy is evidenced by the fact that in the case of *Crisp v Homewood* (1919, 83 J.P. 121) a clerk in the War Office was convicted of betraying an official secret by revealing to a tailor details of contracts between the War Office and contractors for army officers' clothing.

There could be no state secrets in an i.e.p.d. Everyone knows everything, including the fact that nothing so widely known can fail to be known by foreigners.

Further, if one adopts a Marxist interpretation of the nature of the state (and many advocates of the i.e.p.d. do), if, that is, one believes that the state is a committee of the propertied classes the function of which is to maintain a repressive status quo, then in an i.e.p.d. there can be no state, and therefore no state secrets.

It will be objected that even an i.e.p.d. has general interests which can only be protected by secrecy, e.g. war plans in the event of attack. This is obviously correct. But I submit that this is not an argument against an i.e.p.d., merely an argument about an i.e.p.d. "in one country". McMullen has rightly suggested that one human group might be at (say) a military disadvantage if it is an i.e.p.d. when no other group is. The lesson to be learnt is that any movement towards an i.e.p.d. must be world-wide. But this is a matter of politics, of tactics, and has no claim to be discussed in a philosophy journal. McMullen's use of the term 'paradox' suggests that he thinks he has revealed an internal contradiction in the concept of an i.e.p.d. I wish to say that there is a political, but not a philosophical problem here.

*St. David's University College, Lampeter*

## NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, R. G. Swinburne; J. H. Benson, Alec Fisher, Andrew Harrison, R. F. Holland, A. C. MacIntyre, Bernard Mayo, A. R. White, Peter Winch. This committee is responsible for appointing and advising the editor and for the general policy of the paper.

**SUBSCRIPTIONS.** The subscription to ANALYSIS, Vol. 33, is £1.50 for individuals in Great Britain, £1.75 for individuals abroad, and £2.00 for all institutions, post free. Each volume comprises six numbers, five of 32 pages and one of 48 pages, appearing within the academic year—in October, December, January, March, April and June. Orders should be sent to Basil Blackwell, 49 Broad Street, Oxford, or placed with any bookseller.

**CONTRIBUTIONS.** Articles submitted for publication should be addressed to Dr. C. J. F. Williams, 1 Fossefield Road, Midsomer Norton, Bath, BA3 4AS. Contributors are asked to note the following requirements.

Articles should normally not exceed 3,000 words in length. Occasionally, however, longer contributions can be accepted;

They must be typewritten in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

**It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:**

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgement is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope and international reply coupons of the requisite value, whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

The copyright of articles printed in ANALYSIS remains the property of the author, but contributions are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.

ol. 33 No. 5

(New Series No. 155)

April 1972

---

# ANALYSIS

---



Edited by

C. J. F. WILLIAMS

---

## CONTENTS

Ingredients of perception

IRVING THALBERG

Goodness of a kind and goodness from a  
point of view

A. D. M. WALKER

On the evidence of one's "memories"

ANDREW NAYLOR

Nonsense? Nonsense!

HECTOR MONRO

'What a man does he can do?'

BERNARD GERT *and*  
JAMES A. MARTIN

Geach on 'know' in 'if' clauses

D. R. ALTMANN

The incompatibility of omnipotence and  
omniscience

RICHARD R. LA CROIX

---

BASIL BLACKWELL · ALFRED STREET · OXFORD

---

*Forty pence*

## INGREDIENTS OF PERCEPTION\*

*By* IRVING THALBERG

I WANT to elaborate a relatively novel theory of perceiving objects and occurrences. I call it the component analysis. Aside from freshness and simplicity, its virtue is that it provides a niche for three philosophically troublesome kinds of events which are connected with perceiving, and keeps us out of endless wrangling over such events. The controversial goings-on are: (i) electro-chemical happenings in our nervous system when we hear, touch, see, smell and taste; (ii) the putative cavalcade of auditory, tactile, visual, olfactory and gustatory sense data before our minds; (iii) the presence—or “existence”—of a material object, or the unfolding of an occurrence, from which our perception of that object or occurrence is alleged by many philosophers to result.

For brevity, I follow tradition and concentrate on visual perception. I use that term broadly, to cover active scanning, passive gaping, careful and inattentive observing, successful spotting, along with most forms of misperceiving. I must neglect those varieties of perceiving and misperceiving where no publicly visible object or event is on hand. Hence my analysis will not apply straightforwardly to hallucinatory perceiving. Space does not allow me to explain why a percipient is active in some kinds of seeing. I apologize as well for my hasty *résumés* of philosophical debate over neural happenings, sense data, and object-causes of seeing. My goal is to circumnavigate these battles.

## I

Controversy about neural events only heated up during the past century, as physiologists learned that electro-chemical goings-on in our optic nerve, and other regions of our cerebral cortex, are indispensable to sight. When these processes are impeded our visual activities halt. Moreover experimenters can demonstrate that when they subject a person's brain to the right electrical stimulation he will swear that he sees objects and events. Yet no corresponding items are before him. His testimony is less significant, however, than the fact that occurrences in his central nervous system at the time resemble those which take place when suitable objects and events are on view.

Since these brain processes appear to play a leading role in our perceptual undertakings, philosophers have wondered: Is sight perhaps nothing but the firing of neurons? Bertrand Russell's hypotheses were

\* I read an earlier version to the Canadian Philosophical Assn., June 8, 1972, in Montreal. I thank commentator Robert Rose, Professors Braybrooke and Rosenberg, and others, for critical discussion.

more audacious: In *Outline of Philosophy* (London: Allen & Unwin, 1927), he maintained that what you see, and all you ever see, are the sensory terminals of your own brain! He declared: 'All that you see must count as inside your body' (p. 146; compare pp. 131f, 135f, 138f, 140, 144f, 148f). Elsewhere in the same book Russell modified his doctrine to something like this: All you know for sure, when you engage in perceptual adventures, is that some characteristic neural process is occurring. There may be no material object facing you, and your sense organs may not be functioning. But whatever tricks are being played upon you, you can be sure that your neurons are firing. So even if we are not perceptually aware of processes within our skull—since we do not witness them—these are all we have certain knowledge of when we see.

From the opposing side down-to-earth philosophers sometimes take just as paradoxical a view. Exasperated, they deny that neural events have anything to do with the notion of sight. After all, did learned men of antiquity know of such processes? Did they even realize they possessed a cerebral cortex? Yet they, and scientifically unschooled people in more recent times, gained reliable enough information with their eyes about extra-cerebral happenings. More generally, isn't it conceivable that human beings should see but have no nervous system at all, and therefore shelter no electro-chemical activity within their skulls? If so, how can Russell claim to know with certainty of these processes? Logically and epistemically speaking, they seem inessential to perception.

My component analysis of perception takes account of both types of argument. Instead of reducing your visual exploits to neuron firing, we say that this cerebral event is one ingredient of a broader event: your seeing an object outside of your cranium. Two analogies may get across this notion of componency. Think of what happens when an outboard motor runs. The engine is made up of moving parts, *inter alia*. It is not something extra alongside these parts. Its running consists in the sparkplug's firing, the combustion of air and gas inside the cylinder, the movement of pistons, valves and so forth. No sub-event by itself constitutes the engine's operating; nor is the latter an additional occurrence, over and above the sub-events which compose it. Notice also that the component happenings are not mere segments of a more extensive occurrence. Typical segments would be: the first hour of the engine's break-in period, or its operating from the wharf to the jetty. These temporal and spatial slices differ from a component inasmuch as they are homogeneous with the occurrence of which they are fractions. They are operations of the engine, while the sparkplug's firing is not.

My second analogy, from the human sphere, is group behaviour. As a concrete example of collective action, take an urban uprising. The



ingredients of a particular revolt might be: Harry smashing windows, Bill setting fire to the unoccupied Welfare Office, Mary inscribing 'All power to the people!' upon a tenement wall, Sam breaking open the corner grocery, Officer Friendly blasting away with dum-dum bullets, and so on for at least a hundred participants. None of these individual antics will qualify as the urban rebellion. Yet the uprising is not a further happening which mysteriously resulted from, or coincided with, the individual deeds which composed it. Again we can easily distinguish constituent events from temporal and spatial segments of the disorder—for example what happened after National Guard troops arrived, or what developed South of Broadway.

Shortly we shall notice dissimilarities between perceiving, with its neural event component, and these two cases. But we need some terminology. It is obvious that an urban insurrection must encompass some forms of aggressive behaviour by many residents. The concrete acts we listed did not have to occur. Yet if a revolt occurred in Ghettoville then it seems to follow logically that there were some individual deeds of the sort we catalogued. There would have been no revolt unless a minimal number of inhabitants carried out acts which authorities had forbidden them to perform. By contrast, gunfire by policemen like Officer Friendly is not required for there to be an urban insurrection. Thus we shall distinguish between logically *necessary* event components, and *accidental* ones. This does not mean that Officer Friendly must have shot by accident! An accidental component event is simply of a type which need not occur whenever a broader occurrence of a given kind takes place. This distinction is rooted in our event-language alone. Over television, or on the scene, Harry's window-breaking spree will not look to you any more "logically necessary" than the patrolman's gunplay. In fact, the latter event, despite its humble logical status, may be more noteworthy from a "dynamic" or causal standpoint; for it may have either deterred or maddened the insurgents. So we shall label Officer Friendly's shooting a causal ingredient. A component occurrence which is neither logically necessary nor a causal ingredient would be the habitual singing of a solitary rebel: his singing does not result from the behaviour of others around him, and it influences none of his fellow rebels. Finally, we must distinguish between causal components of the uprising, and non-component, perhaps antecedent, occurrences which helped bring about, exacerbate or stop the uprising. The assassination of a community leader may have precipitated, without being a sub-event of, the insurrection. Decisions by faraway non-participants may have had a cooling effect.

One last terminological point: If we turn back to the outboard motor example, we appreciate how arbitrary the "necessary"/"accidental" dichotomy can be. Suppose the motor is running. Does it follow

logically that the engine has a sparkplug, and that the sparkplug ignites periodically? Does combustion have to take place? Must there be any internal goings-on at all for the engine to operate? Perhaps all that is required is the propeller's turning? Evidently we lack a precise definition of what it is for an engine to operate. In this case, then, we waste our time tagging component happenings as either "necessary" or "accidental". On the other hand, ordinary empirical investigation will reveal the degree to which any sub-event is determined by an earlier or simultaneous component, and what influence it has over other component occurrences. I must admit that I have no recipe for picking out events which are ingredients. We do so with ease. It helps if we understand event-words. Common sense and science seem to contribute. But however that may be, we have no difficulty in principle separating causal component events, say of the motor's operation, from the pulling of its starter cable, which merely caused it to begin operating.

Now we can compare these examples with our brain process-perceiving case. The divergence is slight. According to what most speakers mean by 'see', happenings in the percipient's optic nerve and cerebral cortex are not logically necessary ingredients of seeing. But physiologists have demonstrated their causal pre-eminence, from the moment light strikes our retina.

You might admit that changes in our optic nerve result from chemical processes involving the rods and cones of our eye, but ask what effects neural happenings produce. Maybe they trigger sense data. We shall consider that riddle in a moment. Another possibility is that neural events cause beliefs to arise in the perceiver's mind regarding objects and events around him. Such beliefs, incidentally, normally figure as necessary components of perceiving. I say "normally" because I want to leave it open whether there are cases of 'seeing but not believing' and 'seeing but not noticing'. Apart from such questions, we still face a difficulty over belief components. For if mind-body "identity" theorists are right, my believing would be the same event as some process in my brain when I see. Therefore my believing could not be an effect of some brain-process component, and the latter would have no seeing components as effects. Now the hitch, for my analysis, is that if partisans of identity are correct, the same component is and is not logically necessary to seeing. But the contradiction evaporates if we recall that "necessary componenty" is language-dependent. Labelled as an instance of my believing that P, event E is a logically necessary ingredient of my seeing X. Tagged as a neural process, E is not. So if E is both a neural happening and an instance of my believing that P, our only fault is that we have registered a single component twice. Here is a parallel. We are at a committee meeting. Louise is there. We describe her as 'the tallest woman in the room' and as 'the committee chairman'. Under the

first description she is not, while under the second she is necessarily a member of the committee.

I realize that there are many residual obscurities regarding *what* one must believe when one sees an object or event X. I need not resolve these in order to dismiss the debate over perception and cerebral happenings. What matters is that my component analysis will not allow us to hold that neural events *cause* "the perception", or cause us to see X. What goes on beneath our cranium is an ingredient of our seeing X. It makes no more sense to claim that the neural sub-event brought about our seeing than it does to say an individual looter's activity caused the rebellion in which he was participating. What does cause us to see objects and events? That will be our final problem, after we grapple with sense-data.

Before we move on, I ought to summarize how my analysis helps cut off the standard debate. My results are mainly negative, but still therapeutic: visual perception is not reducible to, or identical with, its neural component. The latter may be, or may not be, identical with the percipient's having beliefs about nearby objects and events. That is, my account is compatible with, but does not entail a mental-neural identity thesis. As for brain processes which occur when we see: there is no danger any more that they must deny us access to objects we look at. The neural event will not be what we see, but a vital component of our seeing objects, owing to its importance as an effect of happenings in our eye.

## II

Now we must confront items which physiologists seem barred from examining by microscope or electroencephalogram. Still, it is a natural transition from brain processes to sense impressions. Many philosophers assimilate sense data to the last events which occur in our brain whenever we perceive. Dualistically inclined theorists imagine that the firing of neurons mysteriously coincides with, or perhaps brings about, the intrusion of visual images upon our mind. In the book I cited, Russell himself slides away from his argument that a physiologist with another's brain before him can only peer at the sensory terminals of his own cortex. Soon Russell is maintaining:

In the strict sense, a physiologist cannot observe anything in the other brain he is examining; but only the percepts [sense data] which he himself has when he is suitably related to that brain (*ibid.*, p. 147).

I won't evaluate Russell's or any other writer's methods of proving that we have "percepts", no doubt upon some internal cinema screen—when we are afflicted with after-images, when we undergo optical illusions, when we hallucinate, or maybe whenever we see. Most

beginning philosophy students and laymen with whom I've discussed perception seem predisposed to report items of this kind "in here", entering and leaving their visual field. Why say they cannot have them?

Of course I am tempted to put forth challenges: Can you show me a sense datum; or at least give me some identifying marks so that I will be able to fish one out of my own stream of consciousness? For that matter, where exactly is this stream or field they inhabit? How can you determine when a particular sense datum begins and ceases to exist there? Can it exist outside the stream? Turning to other details, do your visual images have spatial dimensions? Depth? How do you measure these dimensions? What of qualities like weight and temperature? Presumably sense data have colour and shape; but do they come to have these qualities in the way a leaf or wood-carving does? Generally how do you get into contact with a sense datum? Do you see it? With what? Are there optimal conditions for viewing it? Will any circumstances prevent you from getting a complete and accurate look? If some theorist capily rules that we just "sense" a sense datum, I would ask him for instructions in this new activity of sensing. If his next defensive reply is that it is more correct to say we "have" sense data, I would demand an account of this kind of ownership. Some philosophers will majestically brush aside these crude questions, and say that we should regard a sense datum as a "theoretical entity", on a par with electrons and pi-mesons. I admit that it would be inappropriate to make such inquiries about electrons. Although intimidated, I'm still curious. Is there parity between electrons and sense impressions? What cloud chambers, mathematical procedures or crucial observations make it reasonable for you to posit sense data? How can invisible, unsensed and unowned electrons be like visual images, which are somehow known or possessed by one percipient alone?

I pose questions like these because of compassion for anti-sense-datum philosophers, who worry that our sense data will block off "external objects". This would indeed be a consequence if we held that all you are aware of during perception is what crops up in your visual field. A further complaint would be that a sense datum theory makes perception immune from scientific study. Yet I am torn the other way. Simply to deny the existence of sense impressions would, for many ordinary people and venerable philosophers, amount to denying the obvious.

Again my component analysis takes us around the dilemma. If you are deeply attached to your visual field and its flora, keep them. In your case, the event of your receiving an image of a dragonfly will be one constituent of your watching the dragonfly circle overhead. Your scrutinizing, or gaping at, the dragonfly is not reducible to this sense datum component. And it would be a gross *non sequitur* if you concluded

that the only object or event you see—all you are aware of—is this sense datum ingredient. Even if you insist on claiming knowledge of your sense impressions, you may still claim knowledge of the dragonfly in addition. Is it not what you see, however deviously? Perhaps you are one of those individuals who is besieged by sense data when you watch things like this dragonfly. Nevertheless you manage to keep the dragonfly in sight! One last advantage of interpreting sense data as components of perception is that you are free to identify your visual images with brain process components, or to indulge in mental-physical dualism. Moreover, two forms of separatism are possible. You can try to establish causal interaction between brain process and sense datum components. Or you can record regular coincidences between sub-events of these ontologically disparate kinds. If you take an interactionist view, sense data would be causal components. If your approach is parallelist, sense data would have no tie with physical components of seeing, but they would have an effect upon the perceiver's mental state of believing.

If we hark back to the other distinction I made with respect to neural happenings, we may encounter some trouble deciding whether sense data are logically necessary constituents of sight. Presumably not. Many people, like myself, only think they may have sense data when they get after-images or "see stars"—not straightforward cases of seeing an object or event. And most sense-datum philosophers introduce visual images by means of the so-called Argument from Illusion. They write as if it were an empirical discovery, like those about the optic nerve, that sense data blossom in our visual field when we misperceive things, e.g., the oar halfway immersed in water. Yet a few proponents of the Argument from Illusion, and some plain folk, seem to take it for granted, *a priori*, that we have sense data whenever we see. For such thinkers, the issue is only whether we are ever entitled to claim perceptual knowledge of anything beyond our ever-present sense impressions. Here my component analysis simply rules out the "logically necessary" approach to sense data. But isn't such arbitrariness a low price for an analysis which allows the devotee of visual impressions to introduce them, without suffering any of the consequences feared by anti-sense-datum writers?

### III

The last controversial item on my schedule is the object or event we see. According to adherents of the Causal Theory of Perception (CTP), one thing we mean by 'see an object' is that the object figured somehow as a cause of our seeing it. Professor Grice formulates this far more subtly. In his essay which revived debate over the CTP, we read:

[It is] insufficient merely to believe that the perception of a material

object is always to be causally explained by reference to conditions the specification of at least one of which involves a mention of the object perceived. . . . This appears to be a very general contingent proposition; though . . . if the version of the CTP with which I shall be primarily concerned is correct, it (or something like it) will turn out to be a necessary rather than a contingent truth. . . . It may be held that the elucidation of the notion of perceiving a material object will include some reference to the role of the material object perceived in the causal ancestry of the perception (or of the sense impression or sense datum involved in the perception). This contention is central to what I regard as a standard version of the CTP. ('The Causal Theory of Perception', reprinted in G. J. Warnock, ed., *The Philosophy of Perception*, Oxford: O.U.P., 1967, p. 85.)

Nuances notwithstanding, I am unsure what Grice means by 'the perception'. Ultimately he focusses on 'the sense impression or sense datum involved in the perception'. Apparently Grice believes that the English phrase 'It looks to me as if there were an X before me' may serve to report a speaker's visual sense datum. If we reject that questionable assumption, and distinguish between 'It looks to me as if there were an X before me' and 'The object before me looks to me as if it were an X', then we have four versions of the CTP. Deploying my tired dragonfly example again, and speaking unidiomatically, we get:

- (i) I see the dragonfly (circling above me) as a result of its presence (because it is circling above me);
- (ii) The object overhead looks to me as if it were a dragonfly, as a result of its being a dragonfly (because it is one);
- (iii) It looks to me as if there is a dragonfly overhead, and this is a result of the dragonfly's presence overhead;
- (iv) I have a dragonfly sense datum in my visual field, as a result of the dragonfly's aerial activity.

Can we elaborate any of these variants so as to make it clear that the dragonfly's presence brings about my seeing the dragonfly? Seemingly with (iii) and (iv) both in mind, Grice offers a recipe:

the best procedure for the Causal Theorist is to indicate the mode of causal connection by examples; to say that, for an object to be perceived by X, it is sufficient that it should be causally involved in the generation of some sense impression by [*sic*] X in the kind of way in which, for example, when I look at my hand in good light, my hand is causally responsible for its looking to me as if there were a hand before me, or in which . . . (and so on), *whatever that kind of way may be*; and to be enlightened on that question one must have recourse to the specialist. (*Ibid.*, p. 105.)

Whether or not we confound versions (i) through (iv), we shall hear vigorous outcries against the CTP. Here is a sampling: Suppose it is true that when we explain my seeing the dragonfly, the dragonfly should come into the explanation. But don't we beg the question under debate

if we just assume that the explanation here must be causal? What marks it as one? Can we list any well established causal laws relating the presence of such objects to our catching sight of them, or their looking some way to us? Hardly: instead we find laws correlating the presence of such objects and the propagation of light rays from the object, and both with subsequent changes of the perceiver's retina, and so on. When we investigate the causes of perceiving and misperceiving objects, we never learn of the object's presence. Instead we study the favourable and unfavourable influence of illumination, backdrop, distance, camouflage, fog. We establish ideal and hindering conditions of the percipient: previous training, enthusiasm, expectations, myopia, fatigue, suggestibility. If these are the determinants of our perceptual successes and shortcomings, and of how an object looks to us, then versions (i) and (ii) of the CTP lose plausibility. The causes of my seeing the dragonfly, and of the object overhead's looking to me like a dragonfly, do not include the dragonfly itself.

With regard to version (iii), as well as (ii), Professor Alan White airs the following doubts:

[It] is an analytically true statement that the word 'gaol' (a snake, a bush) looks to normal persons in normal conditions as if it were, e.g., the word 'gaol' (a snake, a bush). . . .

. . . Grice's view that the explanation of its here and now looking to me as if there were the word 'gaol' in terms of the presence of the word 'gaol' is of a causal kind would only follow if the non-contingent statement, that its looking to a normal person as if there were an X is in normal conditions due to the presence of an X, expresses a causal law. . . . But . . . I am inclined to say that no statement of a causal law can be non-contingent. Further, is it not the case that, *if* the circumstances are normal, then not only does X look as if it were an X, but also it looks as if there were an X [and] there is an X? (In Warnock, *op. cit.*, pp. 120-121.)

What if circumstances or the perceiver's condition are non-standard? Then those untoward factors would causally account for the dragonfly's looking to me like a desert bat, and for its looking to me as if there is a desert bat overhead. How about the involuted case where I overcome conditions of myself or my surroundings which make dragonflies strongly resemble desert bats? Then whatever enables me to rise above perceptual temptation must be the cause of its looking to me, in spite of everything, as if there is a dragonfly overhead. This vehicle of epistemic grace cannot be the dragonfly itself. We have tried three versions of the CTP, but we have discovered no reason to accept it so far.

A further grumble against interpretation (iii) is that the key phrase 'It looks to me as if there is a dragonfly' is not as a rule used to assert, even guardedly, that one has *seen* a dragonfly. Normally one thereby expresses an estimate or belief, not necessarily based on perceiving the object in question. An example would be: 'It looks to me as if another

monetary crisis has begun'. If defenders of the CTP agree, then variant (iii) is no more than a causal theory of how objects determine one's *beliefs* about one's immediate surroundings. In terms of my component analysis, (iii) records a causal relationship between the object and the cognitive ingredient of my seeing it.

As a causal hypothesis about the dragonfly's effect upon the whole event of my seeing the dragonfly, interpretation (iv) must be disqualified on similar grounds. We noticed earlier why it is implausible to equate the events of perceiving X and having Xish sense data. So at best (iv) illustrates how the sense-datum component of my seeing the dragonfly results from the dragonfly's aerial caper: hardly an instance of X causing perception of X.

Faithful to my pledge of neutrality, I will not attempt to vindicate defenders or attackers of the CTP. From my remarks on variants (iii) and (iv) of that embattled doctrine, it should be obvious how I plan to arrange a truce. It does sufficient honour to the CTP if we say: The dragonfly's presence causes light rays to strike my retina, bringing about chemical transformations in the rods and cones of my eye; this sends electrical impulses along my optic nerve. Further effects would be certain states of my cortex, my new belief that a dragonfly or some insect is overhead, and—if I am afflicted with them—the appearance of a fresh harvest of sense impressions in my visual field. The dragonfly's hovering causes, either proximately or indirectly, each of these effects. Does it bring about the composite event of my seeing the dragonfly? This is not a logical consequence. But if we draw such a conclusion anyway, then we have to deal with the objections, already noted, against 'holistic' readings of the CTP—namely (i), (ii) and perhaps (iii).

On my analysis, the dragonfly's hovering is not just a cause of these other components of seeing the dragonfly. It is an ingredient too. Does that sound outrageous? Not if you reflect on what I'm asserting. To say the dragonfly's presence is part of my seeing is not to say that it is present in me—part of me! That would follow only if you assume, question-beggingly, that the event of my seeing X must somehow be confined to me. But why overstress the possessive pronoun here? Isn't it more correct to hold that N's perceiving X is a relational event involving N and X? Relational occurrences, like X's drifting above N, do not take place "in" X or N. Our feeling of outrage diminishes further if we review what transpires when people are perceptually related to objects. For one thing, it must be the case that X is along N's line of sight. If N has mirrors, the line may be crooked. All this suggests that 'I see the dragonfly overhead' entails 'The dragonfly is overhead'. Its presence is a logically necessary component of my seeing it, as well as a cause of other sub-events.

Scientific and common sense investigations are bound to reveal other



ingredients of perception besides light waves and the three philosophically worrisome elements I have discussed. Historians learn more about what happened during an urban insurrection. Mechanics discover new details of an engine's operation. I have offered no criteria, much less logically necessary or sufficient conditions, for saying what sub-events make up some type of occurrence. Neither do social scientists and garage-men. The risk of proceeding so informally is that I may have no way of excluding any item as a constituent of some event like my perceiving the dragonfly. Without the earth's motion during the past 24 hours, or for that matter the formation of the solar system five billion years ago, would I have been able to see the dragonfly? No; but I sharply distinguished such causes of my current perceptual activity from its causal or non-causal components. So while I admit that I *may* have left the door open for unwanted components, not every causally contributing event will intrude.

Still thinking cosmically, an objector might bring up the stock puzzle of my gazing at an asteroid which exploded millions of years before my birth, although its light is just reaching me now. Did the event of my seeing the asteroid begin so long before I was born? One reply would be that such events take such a long time that they have to start before the lucky observer's birth. But I'm unsure that the question is clear enough to merit a simple answer. What about Einstein on simultaneity? Is it correct to assume that the asteroid blew up at the same time as events which were taking place on earth millions of years ago? A third reply is that the time lag puzzle is no more damaging to my component analysis of perceiving than to any alternative theories.

My goal in propounding this analysis was to clarify the status of neural goings-on, sense data and objects which we feel are somehow connected with perceiving. I've concentrated on avoiding stalemated disputes. Nevertheless my ingredient analysis is compatible with a wide range of metaphysical positions. You can be an identity theorist about sense impressions and beliefs; or you can be a dualist of either parallelist or interactionist leanings. Most important, you can assume that there are objects to be seen, which affect us when we see them.

*University of Illinois at Chicago Circle*

1

## GOODNESS OF A KIND AND GOODNESS FROM A POINT OF VIEW

By A. D. M. WALKER

THE purpose of this note is to outline certain difficulties in J. O. Urmson's recent account in *The Emotive Theory of Ethics* of the distinction between evaluating things as of a kind and evaluating them from a point of view.

Urmson's elucidation of the distinction is by means of examples and by a series of informal characterisations. Thus 'That is a good apple (knife, car)' and 'He (she) is a good golfer (secretary)' are said to exemplify goodness of a kind (p. 99) while 'This road (period of weather) is a good thing from the farmers' point of view', 'The planting of the hard shoulder with flowers would be a good thing aesthetically', 'That is a good plan', 'He acted from a good motive' are offered as instances of goodness from a point of view (pp. 100, 105, 106, 108). And from various remarks scattered throughout Chapter Nine can be collected the three following points of contrast between the two uses:

- (i) in judging good of a kind we assess something as a member of a class that not everything could be a member of, or in a role that not everything could fulfil; whereas in judging good from a point of view we judge from a point of view from which everything could be judged—even if, from that point of view, most things might be judged as indifferent (p. 104).
- (ii) in judging something good of a certain kind the criteria of merit are criteria for that kind of thing; whereas in judging good from a point of view the criteria are used for assessing the merits of things of many different kinds, indeed of any kind (pp. 99–101, 104).
- (iii) in judging good of a kind the relevant criteria are 'determined solely by the character of the object evaluated, by whether it is an apple or golfer or what not that we are judging'; whereas in judging from a point of view 'the relevant standards are determined . . . by the point of view that the judge has adopted, by the interests that he is bearing in mind' (p. 106).

Although Urmson allows that his distinction may not be exhaustive, he does, perhaps with minor qualifications (p. 107), regard the two uses of 'good' as mutually exclusive.

This outline of Urmson's views presents us, I believe, with a serious problem—that of 'harmonising' the examples illustrative of either use with the characterisations contained in the foregoing set of points. For, to put it bluntly, the characterising remarks do not seem always to

match the corresponding examples. Thus (to take (iii) first) are the relevant criteria with goodness of a kind always determined 'solely by the character of the object evaluated'? Are they so determined with, for instance, the goodness of apples? If, as it would seem at first sight, the issue is that debated by Geach and Hare in their papers on 'Good and Evil' (P. T. Geach, 'Good and Evil', *ANALYSIS* 17.2; R. M. Hare, 'Geach: Good and Evil', *ANALYSIS* 17.5), the character of an object determines the criteria of evaluation when, in Geach's formula, given the descriptive force of 'A' the descriptive force of 'a good A' is determinate. But this being so, the criteria for good apples seem *not* to be solely determined by the character of apples, but by our typical interest in them for eating. To be sure, Urmson's claim is more plausible with certain other of his examples—with good knives and good secretaries and in general where 'functional terms', as Hare calls them, are involved (see op. cit. and *The Language of Morals*, pp. 99–103, 133 ff.), though even here qualifications are called for (see *The Language of Morals*, p. 101, and P. R. Foot, 'Goodness and Choice', *Proceedings of the Aristotelian Society*, Supp. Vol. 1961).

Perhaps, however, Urmson would interpret (iii) otherwise, the contrast being rather that in the case of goodness of a kind, though not in the case of goodness from a point of view, the relevant criteria are determined not merely by the *individual judge's* interests or point of view but by something else (in some cases possibly by the *typical* or *standard* interest in that kind of thing). Such an interpretation would of course allow 'That is a good apple' to be retained as a case of goodness of a kind but, unfortunately for Urmson, would require also that many of his examples of goodness from a point of view be similarly classified. When it is said that Smith acted from a good motive or Gabriel was a good angel (p. 109) the criteria of merit are not fixed by the individual speaker's interests or point of view—'good angel' and 'good motive' conventionally presuppose just as much as 'good apple' and 'good knife' the use of certain standard criteria.

Again with (ii), the claim that with goodness of a kind the criteria for assessing something as a good  $C_1$  differ from those for assessing something else as a good  $C_2$  whereas the criteria for assessing merit from a particular point of view apply to items of many kinds, there is difficulty in finding an interpretation which fits the examples. It must first be remarked, parenthetically, that because the two uses of 'good' are not here being compared in parallel respects the characterisation fails to incorporate a genuine contrast. The criteria for something's goodness as a  $C_1$  may well be *applicable* to things of kinds  $C_2$  and  $C_3$ —golfers as well as apples presumably taste or fail to taste sweet. Certainly whether a golfer tastes sweet does not bear on his merit as a golfer; but then a precisely parallel claim can be made, on Urmson's authority, about the

criteria of merit from a point of view: although the criteria of merit from a point of view *apply* to things of kinds  $C_1$ ,  $C_2$  and  $C_3$ , they need have no relevance to the *merits of things as of kind  $C_1$ ,  $C_2$  or  $C_3$* . However, to return to the main theme, is it clear that Urmson's examples, and other obvious examples of each use, in fact match up to these characterisations? Surely, it may be said, should our typical interest in kinds  $C_1$  and  $C_2$  be the same, the criteria of merit as of either of these kinds will be identical; for example, our typical interest in cars, aeroplanes, ships and bicycles is as means of transport, and the criteria of merit as of any of these kinds seem to be safety, speed and comfort.

A natural rejoinder for Urmson is that the relevant criteria must not be conceived in such general terms and that once they are stated with due specificity their apparent identity for merit as of several different kinds will dissolve; thus a good ship should perhaps have a capacity to weather rough seas without pitching and rolling, a condition plainly irrelevant with non-marine forms of transport. However, if to protect his claim about goodness of a kind Urmson requires the criteria of merit to be conceived with such a degree of specificity, his associated claim that in assessing goodness from a particular point of view we always use the *same* criteria becomes markedly implausible: although both a period of sunny weather and a road can be assessed as good from the farmers' point of view—both may advance farmers' interests—the *specific* criteria in terms of which each is so assessed are not obviously identical (the weather may be a good thing from the farmers' point of view because the high temperatures and low rainfall promote crop-growth, the road because its position and construction serve to protect adjacent pastureland from flooding). Or to take another example, we judge both paintings and symphonies to be good from the aesthetic point of view; but is it at all plain that we use in each type of judgment the same specific criteria? Urmson's point (ii), it would seem, can be made to fit his examples only if we trade illegitimately on the elasticity in our conception of criteria of merit.

Finally with point (i) (which arguably also fails to incorporate a genuine contrast) we encounter the same difficulty of reconciling Urmson's examples with his characterising remarks, more especially his claim that the criteria for goodness from any particular point of view are applicable to anything and everything (p. 104).

Of course when Urmson makes this claim in connection with the farmers' point of view he is careful to allow that many things will have no effect whatever on farmers' interests—they will, he says, be judged indifferent from the farmers' point of view. Yet true as this may be, it surely must at least be conceivable, before something can be judged from the farmers' point of view (even if it is then judged indifferent), that it *might* affect farmers' interests (as, for instance (see p. 102), regulations in

the cotton spinning industry might). With the farmers' point of view, admittedly, a background can perhaps always be supplied to make it conceivable that even such implausible candidates as the square root of nine might affect farmers' interests. However, this possibility, if it is one, of assessing everything from the farmers' point of view appears to be a feature not so much of the "point of view" use of 'good' in general as of this particular point of view. Other points of view are, to varying degrees and for various reasons, less catholic in their scope. For example, since necessarily one cannot affect the past, is it conceivable that anything done at the present time may affect the interests of the Greek side at Thermopylae in winning a victory over the Persians? Surely one cannot say that from the point of view of the Greek side at Thermopylae such an event as Britain's entry to the Common Market two and a half millennia later is (or was) an indifferent thing. (Any temptation to hold that this does make sense stems perhaps from confusing the remark with the different one that Britain's entry to the Common Market, had they been able to foresee it, would have been a matter of indifference to the Greeks at Thermopylae. Part of the difference between the two remarks can be captured by saying that something is indifferent from X's point of view if it does not affect X's interests while it may be a matter of indifference to X if, for example, he is not interested in it. Thus something may be a matter of indifference to an individual who has adopted a particular point of view precisely because it cannot be assessed from that point of view. One might suspect, incidentally, that Urmson's claims about the catholic scope of the farmers' point of view derive some of their plausibility from confusing these locutions.) There are, besides, other points of view such as the moral or the aesthetic point of view which appear to belie Urmson's claim. Asked to assess from the moral point of view the boiling point of water, Urmson would presumably maintain that it was morally indifferent. But that would be a clear misuse of the phrase 'morally indifferent', which is reserved for items (like human actions or motives) which are of a kind to be assessed from the moral point of view but happen, from that point of view, to be neither good nor bad; the boiling point of water does not belong to this category. Somewhat similarly (though in some ways differently), the aesthetically indifferent is the aesthetically mediocre; it is again obvious that the boiling point of water cannot naturally be characterised as aesthetically indifferent.

It is, I suppose, unnecessary to insist that the foregoing discrepancies between Urmson's characterising remarks and his illustrative examples obscure whatever distinction is intended by his contrast between goodness of a kind and goodness from a point of view. To suggest that the discrepancies may be ignored because they show merely that Urmson's examples are ill chosen seems implausible; the examples play too large

a part in Urmson's account of the distinction to be so lightly abandoned. On the other hand, I should not wish to deny that within our use of 'good' several independent and overlapping distinctions may usefully be drawn. It may well be that *both* Urmson's set of examples *and* his various characterising remarks indicate significant distinctions; but, if this is so, it is essential that they should not be misrepresented as merely different ways of marking a single distinction.\*

\* I am grateful to G. Wallace for his comments on an earlier draft of this note.

*University of Hull*

## ON THE EVIDENCE OF ONE'S "MEMORIES"

By ANDREW NAYLOR

TO the question 'How is it that we can remember that *p*?' traditional theorists of memory reply that we are able to do this because we have memory-knowledge that *p*, where this is knowledge the certainty of which is grounded upon some present memory-experience, *e.g.* a memory-image. Contemporary theorists reject the latter characterization of memory-knowledge and maintain that grounds for the certainty of our memory-knowledge can lie only in the past. Thus, insofar as the two theories differ over the evidential status of present memory-experience, the traditional theory affirms, whereas the contemporary theory denies, that a person can correctly be described as having *remembered that p*, solely in virtue of his having knowledge that *p* the certainty of which is *grounded* upon his present remembering. My own belief that there cannot be such a case—hereafter referred to as a 'G-case'—derives from what I regard as being the essential content of the contemporary theory: that remembering is basically retention, the continued if not the continuous possession of knowledge (in most cases), or (in other cases) of belief or an inclination to belief, or (as I would like to think) of a certain sort of rational basis for the certainty of any of these.

In his book *Memory* (London, 1971) Don Locke finds much to agree with in the contemporary theory, but holds out nevertheless for a rehabilitated traditional theory of his own, not of memory in general, but of certain personal memory recollections, which he calls 'memories'. 'Personal memory', he says, 'consists in bringing some previously experienced thing to mind, thinking about it again, and going over what it was like, whether we realize that we have actually experienced it or not, where the ability to do this depends on our having experienced it' (p. 76). (In support of this account, he cites considerations brought out in

'Does All Memory Imply Factual Memory?' (ANALYSIS Supplement, 1965)—the authorship of which, incidentally, he attributes not to J. T. Saunders, but to another prominent writer on memory.) The notion of a "memory" is introduced in terms of the following thesis about recalling: where  $x$  is something personally remembered, we can be described as having *recalled*  $x$  only if we remember what  $x$  was like in the sense that we remember the such-and-suchness of  $x$  (p. 80). To remember the such-and-suchness of  $x$  just is, on Locke's view, to have memory-imagery of  $x$ . Such recollections 'are sometimes described as *memories*: a memory is some previously experienced item as we recall it' (p. 81).

One tenet of his theory of "memories" is that "memories" can provide knowledge—in two ways. First, they occasionally are a source of knowledge in that they provide us with the information we remember. Second, they can provide 'evidence in support of what we claim to know' (p. 87). Notice, however, that when this second claim is so stated—in terms of evidence, not for what we claim to remember, but for what we claim to know—nothing is claimed which is contrary to the contemporary theory. What contemporary theorist would not allow that I might just now have come to know that my neighbour is away from home, where this is something I have inferred from the "memory" I just had of newspapers piled on his doorstep? It would seem that if Locke's is to be a traditional theory rehabilitated rather than diluted he must be able to show that "memories" can sometimes provide *the* grounds on the basis of which we are *entitled* to the certainty we would express in claiming to *remember*. He must, in other words, produce a clear instance of a G-case. However, in his entire discussion of the claim that "memories" provide evidence, he at best manages to come up with a case that has all but one of the features of a G-case. This case occurs at the end of the following passage:

(I) . . . although personal memory can provide evidence that what I claim to know is true, it need not be that my own belief is grounded on that evidence. I probably believe it already, without appealing to the evidence of my memories. Or, in other cases, I might be unsure about what happened, and in that case my memories could provide grounds for my own belief: it may be only because I dimly remember, through an alcoholic haze, throwing a bottle and the sound of tinkling glass, that I believe that I broke a window (p. 86).

Although I doubt whether Locke would claim that this is a G-case, there is another case which he would, I think, almost surely want to say is a G-case. The case is one he cites to support the first of the two ways in which he thinks "memories" can provide knowledge—that they can sometimes be a source of knowledge:

(II) . . . someone asks me whether a certain student, whom we both want to see, was present at my lecture. . . . It may be that during the

lecture I did not consciously notice whether anyone was sitting in the back row. But now, when asked, I recall that the student was there, sitting in the back row, though I did not notice it at the time. If I had noticed him, I would have said something to him, but I did not; it is only now that I remember he was there (p. 54).....[I] come to know that he was present, though I did not realize it at the time, because now—and only now, when I think back over the lecture—I can recall him sitting in the back row. My imagery here does not depend on previous knowledge, because I did not know he was there until I had the imagery. Rather, my knowledge depends on memory, in that it is only because I now recall him sitting there that I know that he was present (p. 84).

Locke does take this to be a case in which the knowledge he comes to have is such that he can be described as having remembered that the student was present. And he would, I think, want to claim that the grounds for the certainty of this knowledge are provided by his present “memory”-experience. Indeed, he would have to claim this, or else admit to having produced no counter-example to the contemporary theorist’s position regarding the evidential status of present memory-experience; for, as it happens, he has provided no cases other than those such as (II) which stand a chance of qualifying as G-cases.

I shall now examine case (I). I think that, if I can show why it is that (I) is not a G-case, I can then explain why (II) cannot be one either.

*Case (I).* What “memories” does Locke have in case (I)? Consider first the items experienced: throwing a bottle and the sound of tinkling glass. Since it is as such that he recalls these, there are two “memories” he has, one of throwing a bottle and one of the sound of tinkling glass. Moreover—since whatever else he recalls concerning the incident, he recalls as a specific feature or quality of his bottle throwing or of the sound—these two “memories” exhaust the generic content of what he recalls as he recalls it. These “memories” do account for his coming to believe that he broke a window: upon having them he simply realizes, or perhaps figures out, that this is what he must have done. His belief is, clearly, grounded upon evidence provided by these “memories” alone, but what he believes is, just as clearly, not something he has remembered.

At this point a “rehabilitated” traditional theorist, if he were set upon making (I) out to be a G-case, might say: ‘As Locke has pointed out (pp. 100–101), there are two distinguishable uses of “remember that”—a fact-stating use and a “memory”-describing use. True, it cannot be said, employing the fact-stating use, that Locke “remembers that” he broke a window; nevertheless, this can be said, and would in fact be truly said, employing the “memory”-describing use’.

This move is bound to be abortive. Locke’s distinction between ‘two different, though normally overlapping, ways in which we use [“remember that”]’ is closely tied in his own thinking to a distinction between ‘two kinds of memory-knowledge, between factual memory-



knowledge (knowledge we retain or remember) and personal memory-knowledge (knowledge based, at least in part, on what we personally remember)' (p. 98). 'Based on', as used here, need not mean *grounded on*. It does mean this when personal memory-knowledge is based on "memories" in the way that Locke's certainty that he broke a window is based on evidence provided by his "memories". But there is also, implicit in much of Locke's discussion, another way of thinking about personal memory-knowledge—that, a way of thinking which derives from the idea of having-a-"memory"-of. Thus, when he has the "memory" of throwing a bottle, he will necessarily have had a memory-image of doing this. Memory-imagery is 'a special sort of imagination-awareness of past events themselves' (p. 91). There is within our "memories" a distinguishable element which depends in some way on 'present remembering' (p. 98) and which is 'our knowledge of the things we remember' (p. 92). This element, the personal memory-knowledge, is something over and above 'our memory of facts we already know' (p. 92). Given these claims, it will be natural to think of Locke, when he has his "memory", as having personal memory-knowledge (*i.e.* -awareness) of throwing a bottle, as having this knowledge to whatever extent he is occurrently aware of the particular characteristics, the such-and-suchness, of his own past action. It also will be natural, provided he recalls this action *as* his throwing a bottle, to think of him as having personal memory-knowledge (*i.e.* -awareness) *that* he threw it.

Personal memory-awareness that (*e.g.*) some event,  $x$ , took place may thus be thought of as a limiting case of personal memory-awareness of  $x$ : just as the latter is an image-dependent awareness of  $x$  itself, spread out over the whole face of  $x$ , as it were; the former is the very same sort of awareness (merely) of  $x$ 's having happened. The following relationships may now be noted. (A) Where  $x$  is what you recall *as* you recall it, having a "memory" of  $x$  entails both (1) having some personal memory-awareness of  $x$  and (2) having personal memory-awareness that  $x$  happened. These relationships hold generally, except perhaps for cases where you are unsure or fail to realize that the imagery you are having is a "memory". (B) Where the "memories" you have are of various items each of which is a phase, segment, or characteristic of  $x$ , but you do not recall  $x$  as such, and hence do not recall these items *as* being related to  $x$ , having such "memories", though it may involve (1) as above, logically precludes (2) as above. So if, in a type-B case, you should come to have personal memory-knowledge that  $x$  happened, this would *have* to be knowledge the certainty of which is grounded on evidence provided by your "memories" of other things,  $y$  and  $z$ . And this is how it is with Locke's personal memory-knowledge that he broke a window. In a type-A case on the other hand—I argue for this in the next section—your personal memory-awareness that  $x$  happened *cannot* be knowledge

the certainty of which is grounded on your awareness in memory of  $x$ 's having happened or, for that matter, on anything else in the present. So even if Locke did have such an awareness that he broke a window, it would be 'based on' personal memory only in the sense that it depends on, and in fact consists in, a "memory"-like awareness of having broken it.

I have now distinguished two types of personal memory-knowledge that  $x$  happened or existed: on the one hand, that type the certainty of which is grounded on evidence provided by "memories" of something else,  $y$ ; and on the other hand, that type which is itself a personal memory-awareness-that. Once this distinction is out in the open it is quite obvious why Locke does not, even on the "memory"-describing use, 'remember that' he broke a window. For 'remember that' could be used to describe only *what is there* in his "memories" *to be described*. It is, I would suggest, a requirement for correctly applying 'remember that' in its "memory"-describing use, that it be used to express or report personal memory-awareness-that. But since Locke has no such awareness of having broken a window, he cannot be said to have remembered that he did this.

*Variation on case (I): toward establishing that the certainty of personal memory-awareness-that can be grounded only on something in the past.* Let me now alter the situation described in case (I) by supposing that Locke does have personal memory-awareness of having broken a window, that in having this awareness he is certain that he broke it. Could this certainty be grounded, even in part, upon his "memory"-like awareness of having done this? The issue here is not whether there is any knowledge-that, the certainty of which is not grounded on other knowledge-that; for both traditional and contemporary theorist alike can admit there is such a thing as basic knowledge, *i.e.* knowledge-that- $p$  the certainty of which is justified even though it is not based on any knowledge-that- $q$ , distinct from one's knowledge-that- $p$ . (I shall assume that Locke's personal memory-awareness, as well as any factual memory-knowledge he may have, that he broke a window is basic knowledge.) The issue is whether it can be in virtue of anything in the present that the certainty of such an awareness is justified.

Locke insists that factual memory-knowledge 'is knowledge for which the evidence or grounds lies in the past' (p. 97). But rather than countenancing that this could be how it is with personal memory-awareness too, he seems to assume that anything upon which its certainty is grounded would have to lie in the present. I suspect he may think that this follows from the nature of personal memory-awareness, from the fact (if it is a fact) that a personal memory-awareness-that- $x$ -happened cannot be one amongst several exercises or manifestations of a single piece of knowledge acquired in perceiving or experiencing the occur-

rence of  $x$ . In other words, since such an awareness—that is, according to him, neither a piece of knowledge nor a manifestation of a piece of knowledge, it cannot be retained knowledge; and, he apparently thinks, only when knowledge has been retained can its certainty be grounded on something in the past.

I will for argument's sake allow that Locke's personal memory-awareness that he broke a window is not a manifestation of knowledge which he has retained in the sense that he continues to possess the very same piece of knowledge which he acquired at the time this incident took place. I will not, however, allow that it follows from this that he cannot continue to possess (for the certainty of whatever personal memory-awareness of this kind he may have) the very same justification which he had at the time of the original incident. His justification on that past occasion was simply that he was in a position to know that he had broken a window, e.g. he saw a bottle leave his hand, sail through the air and shatter a window pane. And, I think, on any present occasion when he is aware in memory that he broke it, he is justified in the certainty of such an awareness in no way other than the way he was justified on that past occasion in being sure that he broke it. I think that the certainty of such an awareness is, in other words, justified completely apart from any evidence he may have acquired at a time subsequent to that past occasion; that it is, in still other words, not grounded even in part upon any subsequently acquired evidence, including the evidence (if any) provided by the present awareness itself. For to have basic *memory-knowledge* that  $p$ —whether it be factual memory-knowledge or personal memory-awareness—just is, it seems to me, to have basic knowledge that  $p$  the certainty of which is grounded not in some new way, nor in just any old way, but in no way other than the way it was (or would have been) originally. (I have sought here to extend to basic memory-knowledge one central feature of the type of explication which I tried to work out for certain cases of nonbasic memory-knowledge in "B Remembers that  $P$  from Time  $T$ ", *The Journal of Philosophy*, 1971, pp. 29–41.)

*Case (II).* This is the case involving a "memory" of a certain student sitting in the back row. Allow me to give him the name 'Stu', and to suppose that this is how Locke would refer to him. Now suppose that Locke does, in recalling Stu, have a personal memory-awareness that he was at the lecture. In that case Locke could be said, on the "memory"-describing use, to have 'remembered that' he was present. In fact, as Locke would have it, it is *only* on the "memory"-describing use that this could be said, for he holds that (II) is not a case of factual memory at all (p. 100). Case (II), then, is such that if he were described as having remembered that Stu was present, he could be so described only insofar as his knowledge that this was so is a personal memory-awareness-that.

Could the certainty of such an awareness be grounded upon evidence provided by his "memory" of Stu sitting in the back row? That it could not can be shown as follows.

Locke's "memory" of Stu can plausibly be regarded as made up of two components: one consisting (merely) in the personal memory-awareness *that* Stu was present, the other consisting in the memory-awareness *of* various details and characteristics having to do with the appearance, position in the room, *etc.*, of some student or other (whoever he was). The student, as Locke is aware of him in the first component, is specifically identified, but he is not characterized further; whereas, in the second component, he is not identified specifically, but he is characterized (perhaps even in sufficient detail for his identity to be inferred). Neither of these components could, however, provide the grounds for the certainty of Locke's personal memory-awareness that Stu was present. That the first component could not do so can be shown by reasons of precisely the sort that I gave in the preceding section: since his memory-awareness that Stu was present would obviously be basic memory-knowledge, no part of its certainty could be grounded upon evidence presently obtained, including the evidence (if any) provided by this awareness itself. That the second component could not do so either, can be shown by considerations brought out in discussing case (I). Suppose Locke does have some knowledge the certainty of which is grounded upon evidence provided by this second component. This knowledge would not in that case be personal memory-awareness-*that*, but rather the other of his two kinds of personal memory-knowledge-*that*. Knowledge of the latter sort, as I suggested above, cannot be reported by using 'remember that'. So although Locke may have this sort of knowledge that Stu was present, the fact that he has it in no way goes to show that (II) is a G-case. Moreover, there is no way in which his *entire* "memory" (of Stu sitting in the back row) could provide the grounds for the certainty of his personal memory-awareness-*that*. For these reasons (II) cannot be a G-case.

How is it, then, that (II) can be a case of remembering-*that*? That (II) happens to be a case of coming to have new knowledge does not constitute the insurmountable difficulty which Locke imagines for the contemporary theorist's claim that remembering is basically retention. For this claim can be stated in terms of the continued possession, not of knowledge, but of a particular rational basis for being sure. When understood in this way, the contemporary theory is quite adequate to account for the fact that Locke remembers that Stu was present. In (II), as in my "variation on case (I)", what justifies the certainty of his personal memory-awareness-*that* is a certain situation in the past, *viz.*, that he was then *in a position* to know the truth of what he has presently recalled. He allows that in some sense he *saw* Stu in the lecture audience: '... in

some sense I must have noticed him, perhaps unconsciously, or else I could hardly recall him, as I now do' (p. 55). Although Locke's position at the time of the lecture cannot be described by saying that he *was justified* in actually being sure, it can be described by saying that *had he been sure, and been sure because of what he saw, he would have been justified*. It should now be apparent why my explication of basic *memory-knowledge-that-p* on page 165 refers to the way one's certainty *was* (or *would have been*) grounded originally. For what I want to say—and think the contemporary theorist should say—is that in case (II) Locke's remembering that Stu was present consists in his having basic knowledge that this was so, the certainty of which is grounded in no way other than the way it would have been originally. There is, then, if this explication is correct, a viable alternative to saying that (II) is a G-case.

*Indiana University at South Bend*

## NONSENSE? NONSENSE!

By HECTOR MONRO

It is not as easy as it might seem to discover clear-cut examples of nonsensical attributions; even if we put all English adjectives in one hat, and all nouns in the other, and drew them out at random, we would find that the strangest combinations yield possible meanings upon reflection . . . someone may find a meaning in it that we have overlooked, perhaps by inventing a line of poetry in which the attribution works. . . . I should think 'participial biped' and 'a man in the key of A flat' are very probably incapable of being explicated and also such old philosophical favourites as 'consanguinity drinks procrastination'. M. C. Beardsley, *Aesthetics*, p. 143.

A T family reunions  
 (Weddings, funerals,  
 The joint junketings of Christ and Saturn)  
 Blood is so much thinner than whisky and water  
 That consanguinity drinks procrastination,  
 Postponing the ineluctable anacoluthon  
 In the polished Jamesian discourse of Uncle Fred  
 (That prosy participial biped)  
 When he and Uncle Arthur  
 (Literal, inarticulate,  
 A man transposed wholly into the key of 'A flat')  
 Discover  
 That there are impediments to the marriages of mind  
 Certainly more than kin, as certainly less than kind.

*Monash University*

## 'WHAT A MAN DOES HE CAN DO'?

By BERNARD GERT and JAMES A. MARTIN

IT seems to be a commonly held view that what a man does, he can do; that if a man does X, then he can do X. But, in fact, as will be shown here, on one common and important interpretation this view is false; although statements of the 'does' implies 'can' view often permit several different interpretations, some of which are true, at least one is not. In *Mental Acts* Geach provides us with a statement of the view that 'does' implies 'can' which is intended in just the way that is false, as we will show. Here is what Geach says (p. 15).

... to say that a man has a certain concept is to say that he *can* perform, because he sometimes *does* perform, mental exercises of a specifiable sort. This way of using the modal word "can" is a minimal use, confined to a region where the logic of the word is as clear as possible. *Ab esse ad posse valet consequentia*—what is can be, what a man does he can do; that is clear if anything in modal logic is clear, and no more than this is involved in my talking of concepts.

Before we show that as Geach intends it, the "'does' implies 'can'" view is false, let us discuss some ambiguities of its statement. To facilitate discussion, we will take the following expression of the view in question as its official statement.

(P) If S *does* X, then S *can* do X.

There are two sets of ambiguities: there are different senses of 'can' and different senses of 'if ... then ...'

It is widely recognized that one can distinguish a very great number of senses of 'can'. (See e.g. 'Voluntary Abilities' by T. Duggan and B. Gert in *The Nature of Human Action*, edited by Myles Brand, and Bruce Aune's article 'Can' in *The Encyclopedia of Philosophy*.) For example, there is the 'can' of logical possibility. To say that S does X is not self-contradictory is to say that it is logically possible for S to do X. This sense of 'can' we will call 'L-can'. There are indefinitely many other, stronger senses of 'can' which one might separate out here, e.g., philosophers often speak of the 'can' of physical possibility, which implies only that the performance of an action is not contrary to the laws of nature; but the sort of 'can' which primarily interests us here has to do with a person's physical and mental abilities. We call this the 'can' of abilities or 'A-can'.

All competent tennis players A-can hit a tennis ball, and they A-can serve. But only in some weaker sense of 'can' can all such persons hit the boundary lines of the area into which they serve. Only a very exceptional player A-can do this. (Perhaps many tennis players have whatever

abilities are required to *learn to* hit the line on serve; but rather fewer of them A-can do it.) One would be disappointed (or perhaps relieved) to find that his opponent, highly touted as a man who can hit the line, can do this only in the weaker sense of being able to serve the ball and have it sometimes, rather by chance, hit the line. Or suppose that, being told of a man who can distinguish by taste several similar scotch whiskys, we test him and discover that as chance would have it, but no more often than that, he does sometimes produce the correct answers. We will be disappointed because we will have been led to expect a man who gets some large percentage of the answers right.

Although it is not our purpose here to present and defend an analysis of 'A-can', what we have to say will be easier to follow if the analysis we believe to be correct is before us. To say of a man that he A-can do X is to say that, if he wills to do X, then simply given a reasonable opportunity, he will do X a reasonable number of times. (Problems concerning the gaining and losing of abilities, temporal reference, and so on, will not be dealt with here. For a more detailed discussion of the analysis we use see 'Voluntary Abilities'.) What constitutes a reasonable opportunity and a reasonable number of times quite properly depends upon various details of the case in question, e.g. the sort of action X is.

As with 'can', there are several senses to 'if . . . then' or implication, but two will do for our purposes. Again we start with a strong sense: entailment. One might mean by 'if A then B' that the conjunction of 'A' and 'not-B' is logically inconsistent; that if it is true that A, then it is true that B no matter what else, so to speak. Let's say in such cases that 'A' L-implies 'B'. On the other hand there is room for a weaker kind of 'if . . . then' which has to do with the giving of good evidence, or with our being justified in asserting B, given that A. One might say that 'A' implies 'B' in this sense when the truth of A normally entitles us to believe that B, i.e., when A is something which in normal circumstances constitutes good evidence for B. Of course there is room, we suspect, for lots of importantly different kinds of implication within this rough description (this is doubtless true of our senses of 'can' also), but enough has been said to show that this sense (or these senses) is clearly different from entailment. And enough too to show that it is a weaker sense, since A can provide support for B in this sense without entailing it. We might call this the evidential sense of 'if . . . then' and say of A and B in such cases that 'A' E-implies 'B'.

Now we can say with some precision just what version of the 'does' implies 'can' principle is false, that is, what interpretation of (P) is false. The following principle—call it (P\*)—is false.

(P\*) 'S does X' L-implies 'S A-can do X'.

Yet it is just this version which is often assumed to be true, and which Geach employs.

Let us be clear about the falsity of (P\*) first. The man who occasionally hits the service line in the example offered above does not have the physical ability to do so. Suppose that out of one hundred attempts to do so, under reasonable circumstances, he succeeds in hitting the line twice. It is clearly not entailed by his doing this (hitting the line twice) that he A-can hit the line. Indeed, it is strongly suggested by the complete description of the situation that it is false that he A-can do this; that he has not got the physical ability to do it. Take a simpler example. Sometimes when a man flips a coin over three feet, it comes up heads—it lands with the head side up. But it obviously doesn't follow from this that he has the ability to flip heads. In fact, if we know that he is trying, that nothing is interfering with any relevant abilities he might have, his getting heads hundreds of times, but not much more than around fifty per cent of the times he tosses the coin, will be part of the evidence that he has *not* got the physical ability to flip heads. (It is worth noticing too that what we might call the inverse of (P\*), that failing to do A entails that one does not have the ability to do A, is also false though sometimes accepted (see Aune, *op. cit.*). E.g., if he failed to flip heads only twice in a thousand attempts, this does not show that he A-cannot flip heads.)

This point holds as well for "mental" abilities. But with mental abilities another factor enters into the situation. Most, if not all, mental activities are such that we do not say of anyone that he is performing that mental activity unless we believe that he has the mental ability to perform that activity.<sup>1</sup> E.g., we do not say of someone that he is reading unless we believe that he has the ability to read. Thus when we see someone looking at a book and uttering sounds which by an established linguistic convention can be correlated with the marks printed on the page, we do not describe such a person as reading unless we believe that he will be able to utter the appropriate sounds when looking at a number of different books. Now if Geach is maintaining that 'it follows from the fact that a person is reading that he has the ability to read', he is certainly correct, but this is not a fact of modal logic, but is a logical feature of the concept of reading. In fact, such a feature may provide a theoretical basis for a distinction between mental and physical activities. Thus it would be pointless for Geach to hold that reading entails the ability to read, for he is trying to show that simply from observing someone doing something we can infer that he has the ability to do that thing. But if we can't know whether what we are observing is a case of doing that thing (e.g., reading) unless we already know that the person has the ability, we are involved in a rather small circle. Thus clearly the kind of action that Geach refers to when he says 'what a man does' must be

<sup>1</sup> We are indebted to our friends William Ruddick and David Welker for suggesting this point.



describable in a way that does not presuppose the ability to do that thing.

In a way all this seems quite strange. How, one wants to ask, *can* it be the case that a man ever does something that he *cannot* do? Merely pointing out that there are several senses of 'can' is only a bare beginning in answering this question. But the examples given above make it obvious that S's doing X does not entail that S A-can do X, that S has the physical or mental ability to do X. It should be clear enough now that the relation between 'does' and 'A-can' is not entailment, or what we have called L-implication. Could it, then, be E-implication? Not in any simple way: Whether successful performance of X E-implies that one A-can do X depends not only on the circumstances in which one does X, but also on the kind of action X is. If X is an extraordinarily difficult or complex action, e.g., swimming the English Channel, then doing it just once certainly E-implies that one A-can do it. But if it is a simple task, one that could easily occur by luck, e.g., flipping a coin that lands heads, doing it once does not E-implicate that one A-can do it.

Now we have maintained that Geach means (P) in the sense of (P\*); that he claims that 'does' entails the 'can' of ability. But it is not, as can be seen from the above remark, so important that entailment be precisely what is meant. There is no doubt from what Geach says that he means entailment, but in fact even the evidential 'if . . . then' usually will require more in its antecedent than one successful performance. Geach says 'what is can be, what a man does he can do; that is clear if anything in modal logic is clear'. What is clear in modal logic is, of course, that '*p*' entails 'It is possible that *p*'. Geach goes on to say that 'no more than this (the principle of modal logic) is involved in my talking about concepts'. It seems clear then from this passage that Geach thinks that 'does' entails the requisite sense of 'can'. His mistake here lies in failing to distinguish *logical possibility* or 'L-can', from *physical* or *mental abilities* or 'A-can'. For it seems clear when we focus on what he wants to do with the 'can' that allegedly follows from a successful performance, we will see that the 'can' required is 'A-can', the 'can' of abilities, not 'L-can'.

Though Geach does not, in the passage cited, use the word 'ability', the whole thrust of the chapter in which this passage occurs is precisely that having concepts is just having certain sorts of abilities. Indeed, in his summary of this chapter (chapter five) in the table of contents, he says 'Concepts are for me specific mental abilities'. We can see from this that what Geach means in his talk of 'can' must be the 'can' of ability. But, of course, to say of a person that he L-can do X is not to say that he has the mental ability to do X.

It is clear enough that Geach does mean the 'can' of abilities, but perhaps it would be helpful to say something about why he needs this sort of 'can'. He wants to show that having a concept is a matter of

having certain abilities; that a person's having a concept is just a matter of his being able to do certain kinds of things. There are several senses in which a person may be said to be able to do a thing, several sorts of 'can'. But most of them are too weak to be used in attributing abilities; and it is clear that a very strong sort will be required to make plausible the claim that a person's having a concept is constituted by the fact that he *can* do various things, that he is able to do them. But it is also clear that no such sufficiently strong sense of 'can' will be entailed by one or some successful performances (unless, of course, we don't count an action as a genuine performance, if the person does not have the ability). Whether it will be E-implied by the latter depends on such things as the sort of act in question, the proportion of successful performances within the total number of attempts, and whether or not the person in question wants to succeed in the particular situation.

As noted above, there may very well be activities which are so complex or difficult that their performance even once is sufficient to show (E-imply) that one has the appropriate mental or physical ability. But we see no reason to suppose that all the abilities which might plausibly be maintained to constitute concepts should have activities of this sort associated with them. For example, one ability which might typically partly constitute having the concept of a heavy scotch whisky would be the ability correctly to answer the question after sipping a scotch, 'Heavy, or light?' Yet the successful performance of this "activity" (e.g., successfully answering one such question) could hardly be thought to *show* that one has that ability. However, Geach might have in mind only abilities such as those whose complexity and difficulty are such as to make even one successful performance very good evidence. One cannot tell from reading Geach, however, for no detailed account of the sorts of activities he has in mind is provided. He seems to have in mind such activities as using some appropriate bit of language, but a very great deal more than this would have to be said in order to legitimize the jump from 'doing' to 'can do'.

Geach holds the following view.

- (1) S has the concept C if, and only if, S has some ability A.

The following four theses would support Geach's view that if a man 'sometimes *does* perform mental exercises of a specifiable sort' then he has a concept.

- (2) That S does X L-implies that S A-can do X.
- (3) That S does X E-implies that S A-can do X.
- (4) That S uses language in some very complex way (utters sounds to which linguistic conventions apply, etc.) L-implies that S A-can use language in that way.

- (5) That S uses language in some very complex way (utters sounds)  
E-implies that S A-can use language in that way.

We have raised no questions here concerning (1). What we hope to have shown in this paper is that (2), (3) and (4) are false. Geach probably held (2) because he failed to distinguish between L-can and A-can, that is, between logical possibility and mental or physical ability. The falsity of (3) is evident from the fact that one can present many cases where S's doing X does little or nothing to show that S has the ability to do X. (4) may be true if 'using language in some complex way' is taken as *presupposing* the ability to use language, but then it would be trivial. And (5) may be true, but Geach has not provided a detailed enough account to say for sure.

The most important point is thus not directed at Geach's view of concepts (as expressed in (1)), but rather at the general sort of argument he uses to support it: he seems to suppose that since (2) is true—and we have shown that it is not—(4) must be true. We think it is important to note that when speaking of abilities not only is the statement (P) 'If S does X, then S can do X' not a truth of modal logic, it is not a truth at all. Indeed, it is not a *general* truth even when 'if . . . then' is taken in a weaker sense than that of logical implication; that is, even if (5) is true, (3) is not.

*Dartmouth College*

*Professor Geach writes:*

If a man does perform a feat, that shows conclusively that he can perform it, that the feat is within his range; it does not merely show that his having performed the feat is logically possible (and it is quite gratuitous on the authors' part to conjecture that I may have been confusing capacities with logical possibilities). If a man has once done a thing, he *could* do it, however much this may surprise him and others, and whether or not he can do the like again if he tries; anyone who bet beforehand that the man couldn't do the thing has lost his money. There is of course an important distinction between what a man is shown able to do when he does it by a fluke, or again under extreme pressure like a threat of execution, and what a man can regularly do. But this is doubtfully relevant to my discussion in *Mental Acts* of concepts as mental capacities; it has (to put it mildly) no clear sense to say of a man 'It was only by a fluke, or under a grievous threat, that he performed that judgment—he had not the *concepts* that would normally be needed even to think of the matter'.

*University of Leeds*

## GEACH ON 'KNOW' IN 'IF' CLAUSES

By D. R. ALTMANN

WHILE discussing the topics of assertion and "assertoric force", Professor Geach argues that when 'A knows that  $p$ ' occurs in a conditional statement the speaker is not committed to asserting (or assenting to the truth of) ' $p$ '.<sup>1</sup>

The example given by Geach as establishing this is indeed an instance where the speaker would not be committed in this way. But it is not clear that Geach can argue from his example to the conclusion that *whenever* 'A knows that  $p$ ' occurs in a conditional, there is no commitment to asserting ' $p$ '. And Geach does intend a strong claim of this sort, for he contrasts the conditional behaviour of 'A knows that  $p$ ' with that of 'A fancies that  $p$ ', 'A is under the illusion that  $p$ ', finding a commitment to ' $\sim p$ ' in the latter two cases to which there does not correspond any commitment to ' $p$ ' in the former.

'A fancies that  $p$ ' is exponible, according to Geach, into the pair of assertions 'A thinks that  $p$ ' and 'It is not the case that  $p$ '; 'A is under the illusion that  $p$ ' is similarly exponible into 'A is under the impression that  $p$ ', and 'It is not the case that  $p$ '. But they are not logically equivalent to assertions of conjunctive propositions, as Geach indicates (*ibid.*): 'Assertions thus exponible will certainly retain part of their assertoric force when put, for example, into an "if" clause; thus, one who asserts, "If A is under the illusion that  $p$ , then  $q$ ", does not mean "If A is under the impression that  $p$ , but it is not the case that  $p$ , then  $q$ "—rather, he both asserts, "It is not the case that  $p$ " and asserts, "If A is under the impression that  $p$ , then  $q$ ".' Geach then continues: 'Notice that no such complication arises for the verb "know". Use of the expression "... knows that  $p$ " does not commit the speaker to asserting " $p$ "; to adapt an example of Hintikka's, one who asserted in 1916, "If Russell knows that Wittgenstein is dead, then Wittgenstein is dead" would not himself be asserting, "Wittgenstein is dead".'

It is important to emphasise here the distinction between cases where ' $p$ ' logically implies ' $q$ ' (where ' $\sim q \supset \sim p$ ' is true), and cases where the use of an expression in ' $p$ ' commits one to asserting ' $q$ ' (where ' $\sim q \supset \sim p$ ' may be false or ill-formed). First, the use of 'fancies that', 'is under the illusion that' in 'Does A fancy that  $p$ ?', 'Is A under the illusion that  $p$ ?', commits one to asserting ' $\sim p$ '; but there is quite unequivocally no logical implication.<sup>2</sup> Secondly, letting ' $Fap$ ' stand for 'A fancies

<sup>1</sup> In 'Assertion', *Philosophical Review*, 74, 1965, p. 456.

<sup>2</sup> A point of this sort is made by P. F. Strawson in 'A Reply to Mr. Sellars', *Philosophical Review*, 63, 1954, especially pp. 217-8.

that  $p$ , ' $F'ap$ ' for 'A thinks that  $p$ ', both ' $Fap \supset q$ ' and ' $\sim(Fap \supset q)$ ' commit one to asserting ' $\sim p$ ' (with ' $F'ap \supset q$ ', ' $\sim(F'ap \supset q)$ ' respectively). ' $(\sim \sim p) \supset \sim(Fap \supset q)$ ' may be false, and again there is no logical implication of ' $\sim p$ '.

Following Geach, then, one who asserts 'If Russell fancies that Wittgenstein is dead, then I will ring him', is committed to asserting 'Wittgenstein isn't dead' and 'If Russell thinks that Wittgenstein is dead, then I will ring him'. Geach's example with 'know', however, is not of the same form. If ' $Kap$ ' stands for 'A knows that  $p$ ', and ' $K'ap$ ' for 'A "believes" that  $p$ ',<sup>1</sup> then Geach's example indicates that ' $Kap \supset p$ ' does not commit one to asserting ' $p$ ' and ' $K'ap \supset p$ '. In this case the assertion that is supposed to be cancelled by the conditional, ' $p$ ', occurs in it as the consequent. It is this occurrence of ' $p$ ' in the 'then' clause that is responsible for the absence of a commitment to ' $p$ ' in Geach's example, rather than any dissimilarity between 'know' and 'fancy'. For if ' $Kap \supset p$ ' carries no commitment to ' $p$ ', this could either result from the failure of both ' $X \supset p$ ' and ' $Kap \supset q$ ' to carry it; or just from failure in ' $X \supset p$ ', while in general ' $Kap \supset q$ ' does commit one to asserting ' $p$ '. And it is clear that the latter is the case, given that in general ' $Kap \supset q$ ' does carry a commitment to ' $p$ ', and that ' $Fap \supset \sim p$ ' does not: 'If Russell knows that Wittgenstein is dead, then I will ring him' does carry a commitment to asserting 'Wittgenstein is dead', while 'If Russell fancies that Wittgenstein is dead, then Wittgenstein isn't dead' does not commit one to asserting 'Wittgenstein isn't dead'. In general, even though 'X' might be such that ' $X \supset q$ ' commits one to asserting ' $p$ ' and ' $X' \supset q$ ', ' $X \supset p$ ' will not.<sup>2</sup> So Geach's example fails to show that no 'complication' arises for the verb 'know'.

<sup>1</sup> ' $K'ap$ ', or 'A "believes" that  $p$ ' is to be taken as some necessary condition for the truth of ' $Kap$ ' other than ' $p$ '. This is not to say that ' $p$ ', ' $K'ap$ ' are together sufficient for the truth of ' $Kap$ '. (It is not problematic to say that ' $Kap$ ' logically implies ' $p$ ' and ' $K'ap$ ', or that ' $Fap$ ' logically implies ' $\sim p$ ' and ' $F'ap$ ', so long as one can allow commitment from the use of an expression, and logical implication to co-exist—and I can see no reason why one can't. In ' $Kap \supset q$ ', ' $Fap \supset q$ ', however, the logical implication is cancelled, and *only* the commitment to ' $p$ ', ' $\sim p$ ' carries through. This part of the "behaviour" of ' $Kap$ ', ' $Fap$ ' cannot be simulated by an ordinary conjunctive assertion).

<sup>2</sup> It might appear that in general when 'X' is in a context which involves its explication, then the normal commitment is absent. For example, 'If the present king of France is bald, then there is a present king of France' does not commit one to asserting 'There is a present king of France'. Commitment is not consistently absent, however, in these cases: cf. ' $Fap \supset \sim Kap$ ', ' $Fap \supset (\exists x)(F \times p)$ ', ' $Fap \supset (\exists q)(Faq)$ ', and the corresponding statements with 'know'.

## THE INCOMPATIBILITY OF OMNIPOTENCE AND OMNISCIENCE

By RICHARD R. LA CROIX

CONSIDER a finite (not omnipotent and not omniscient) being *b* and an act *a* such that *b* has as some of its self-consistent properties the properties ascribed in the statements:

- (1) *b* does *a*
- (2) *b* is the *only* being who knows that (1) is true.

Now, if '*x* is omnipotent' entails

- (3) *x* can create any finite being *y*, provided (i) *y* has properties such that the statement that *y* has those properties neither is self-contradictory nor entails a contradiction and (ii) there is no being *z* such that the statement that *z* and *y* exist either is self-contradictory or entails a contradiction,

then an omnipotent being can create *b* because the statement that *b* has the properties described in (1) and (2) does not appear to be self-contradictory or to entail a contradiction, and *a* can be an act and *b* a being such that the existence of *b* is consistent with the existence of all other existing beings. An omnipotent being *can* create a being who performs an act known only to himself and to no other being.

But, if '*x* is omniscient' entails

- (4) For any finite being *B* and for any act *A*, if *B* does *A* then *x* knows that *B* does *A*,

then an omniscient being cannot create *b*. If an omniscient being created *b* then by (1) *b* does *a* and by (2) the omniscient being does *not* know that *b* does *a* and is, hence, not omniscient. An omniscient being *cannot* create a being who performs an act known only to himself and to no other being.

Since an omnipotent being *can* create *b* and an omniscient being *cannot* create *b* then '*x* is omnipotent and omniscient' entails

- (5) *x* can create *b* and *x* cannot create *b*.

But, since (5) is self-contradictory it would appear to follow that a being who is both omnipotent and omniscient is logically impossible.

*Detroit, Michigan*

## NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, R. G. Swinburne; J. H. Benson, Alec Fisher, Andrew Harrison, R. F. Holland, A. C. MacIntyre, Bernard Mayo, A. R. White, Peter Winch. This committee is responsible for appointing and advising the editor and for the general policy of the paper.

**SUBSCRIPTIONS.** The subscription to ANALYSIS, Vol. 33, is £1.50 for individuals in Great Britain, £1.75 for individuals abroad, and £2.00 for all institutions, post free. Each volume comprises six numbers, five of 32 pages and one of 48 pages, appearing within the academic year—in October, December, January, March, April and June. Orders should be sent to Basil Blackwell, 108 Cowley Road, Oxford, or placed with any bookseller.

**CONTRIBUTIONS.** Articles submitted for publication should be addressed to Dr. C. J. F. Williams, 1 Fossefield Road, Midsomer Norton, Bath, BA3 4AS. Contributors are asked to note the following requirements.

Articles should **normally** not exceed 3,000 words in length. **Occasionally**, however, longer contributions can be accepted;

They must be typewritten in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:

Authors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgement is required, a return address should also be enclosed.

Authors who wish to have their MSS. returned should enclose a stamped addressed envelope with a return address and a postal reply coupon of the requisite value.

Manuscripts will be sent to authors for correction, and typescripts will be retained by the editor. Authors who have kept their own copies.

Manuscripts remain the property of the author. In their own interest, to avoid loss of their articles.

Vol. 33 No. 6

(New Series No. 156)

June 1973

# ANALYSIS

Edited by

C. J. F. WILLIAMS

---

## CONTENTS

Reverse discrimination and compensatory  
justice

PAUL W. TAYLOR

Reparations to wronged groups

MICHAEL D. BAYLES

Individuals, groups and inverse discrimination

ROGER A. SHINER

Outcomes and abilities

BERNARD GERT *and*  
JAMES A. MARTIN

Justice in a trial

MICHAEL LEAHY

On undetectable differences in sensations

CARROLL LEWIS

Underdetermination of theory and indeterminacy of translation

R. KIRK

Wishful thinking and self-deception

BÉLA SZABADOS

Russell on classes as logical fictions

STEPHEN E. BÖER

---

BASIL BLACKWELL · ALFRED STREET · OXFORD

---

*Forty pence*



# REVERSE DISCRIMINATION AND COMPENSATORY JUSTICE

By PAUL W. TAYLOR



TWO articles have recently appeared in ANALYSIS concerning the apparent contradiction between:

- (1) At time  $t_1$  members of group G have been discriminated against on the basis of a *morally irrelevant* characteristic C

and:

- (2) At time  $t_2$  characteristic C is a *morally relevant* ground for making reparation to members of group G.

In the first article ('Discrimination and Morally Relevant Characteristics', ANALYSIS 32. 4, pp. 113-114) J. W. Nickel presents what he calls the 'reverse discrimination argument' and offers a counter-argument to it. The 'reverse discrimination argument' is that, if we grant as a matter of compensatory justice special advantages or benefits to persons who have been unjustly treated on the basis of a morally irrelevant characteristic (such as being a woman, being black, being a Jew, etc.), we are in effect using a morally irrelevant characteristic as if it were morally relevant and thus still engaging in an unjust treatment of persons. Hence if we are to be just we must avoid reverse discrimination. Nickel's counter-argument is that the special treatment given to persons having characteristic C at time  $t_2$  is not grounded on the (morally irrelevant) characteristic C, but on the (morally relevant) characteristic C', namely, being a person who has been discriminated against because he was C.

In an article by J. L. Cowan ('Inverse Discrimination', ANALYSIS 33. 1, pp. 10-12) Nickel is criticized for failing to realize that characteristic C' is actually a complex predicate made up of a conjunction of two characteristics—C", namely: having been discriminated against (for whatever reason), and the characteristic C itself. Cowan's point is that C" is a morally relevant characteristic in matters of compensatory justice but that C remains morally irrelevant. He holds that reverse discrimination is wrong, since it involves giving people favourable treatment *because they are C* and hence using a morally irrelevant characteristic as a justifying ground for special treatment. Nevertheless, he concludes, special treatment should be extended to any *individual* who has been discriminated against (for whatever reason) in the past, that is, anyone who has characteristic C". For it is a requirement of compensatory

justice that reparation be made to those who have been dealt with unjustly (for whatever reason).

In this paper I want to defend three views, all of which are inconsistent with the claims made by Nickel and Cowan. (I) With respect to the principle of compensatory justice, characteristic C has been *made* a morally relevant characteristic by those who engaged in a social practice which discriminated against persons because they were C. (II) Since C is a morally relevant characteristic at time  $t_2$  with respect to the principle of compensatory justice, that principle requires reverse discrimination. (III) The reverse discrimination in question is aimed at correcting an injustice perpetrated at time  $t_1$  by a social practice of discriminating against C-persons because they were C. Given this aim, the reverse discrimination must be directed toward the class of C-persons as such. Furthermore, the obligation to compensate for the past injustice does not fall upon any particular individual but upon the society as a whole. The society is obliged to establish a social practice of reverse discrimination in favour of C-persons. (It is assumed, of course, that this practice will be consistent with all other principles of justice that may apply to the action-types which are involved in carrying it out.) I offer the following considerations in support of these views.

# I

Suppose there is a socially established practice at time  $t_1$  of unjustly treating any person who has characteristic C, such treatment being either permitted or required on the ground that the person is C. For the purposes of this account I hold that the treatment in question is unjust because characteristic C would not be mutually acknowledged as a proper ground for such treatment by all who understood the practice and took an impartial view of it (in accordance with John Rawls' *A Theory of Justice*, Harvard University Press, 1971).

When a social practice of this kind is engaged in, the members of the class of C-persons are being discriminated against because they are C. By reference to the rules of this practice, having C is a relevant reason or ground for performing a certain kind of action which is in fact unjust (though not recognized to be so by the practice itself). In this context the characteristic C is not accidentally or contingently associated with the unjust treatment, but is essentially tied to it. For the injustices done to a person are based on the fact that he has characteristic C. His being C is, other things being equal, a sufficient condition for the permissibility of treating him in the given manner. Within the framework of the social practice at  $t_1$ , that someone is C is a ground for acting in a certain way toward him. Therefore C is a relevant characteristic of a person.

But is it *morally* relevant? The answer to this question, it seems to me, is that at time  $t_1$  characteristic C is not morally relevant, but, if we

accept the principle of compensatory justice, at time  $t_2$  it is. The principle of compensatory justice is that, in order to restore the balance of justice when an injustice has been committed to a group of persons, some form of compensation or reparation must be made to that group. Thus if there has been an established social practice (as distinct from an individual's action) of treating any member of a certain class of persons in a certain way on the ground that they have characteristic C and if this practice has involved the doing of an injustice to C-persons, then the principle of compensatory justice requires that C-persons as such be compensated in some way. Characteristic C, in other words, has become at time  $t_2$  a characteristic whose *moral* relevance is entailed by the principle of compensatory justice. In this kind of situation, to ignore the fact that a person is C would be to ignore the fact that there had been a social practice in which unjust actions were directed toward C-persons as such.

## II

Given that characteristic C is morally relevant to how C-persons are to be treated if compensatory justice is to be done to them, it follows that reverse discrimination is justified. For this is simply the policy of extending special benefits, opportunities, or advantages to the class of C-persons as such. Contrary to what Nickel affirms, this is not selecting C-persons for special treatment on the basis of the complex characteristic C', namely: being a person who was discriminated against because he was C. For even if the individual C-person who now enjoys the favourable compensatory treatment was not himself one of those who suffered injustice as a result of the past social practice, he nevertheless has a right (based on his being a member of the class of C-persons) to receive the benefits extended to all C-persons as such. This follows from our premise that the policy of reverse discrimination, directed toward anyone who is C because he is C, is justified by the principle of compensatory justice.

Cowan claims that compensatory justice does not require a policy of treating all C-persons favourably (other things being equal) because they are C. His argument is that, if the original unjust treatment of C-persons was unjust precisely because their being C was morally irrelevant, then 'there is and can be no morally relevant group which could have suffered or to which retribution could now be made' (p. 11). My reply to this argument is that the moral relevance or irrelevance of a characteristic is not something that can be determined outside the framework of a set of moral principles. It is true that the principles of *distributive* justice were transgressed by the past treatment of C-persons precisely because, according to those principles, characteristic C is morally irrelevant as a ground for treating persons in a certain way. Nevertheless, according to

the principle of *compensatory* justice (which applies only where a violation of other forms of justice has taken place), the fact that systematic injustice was directed toward a class of persons as being C-persons establishes characteristic C as morally relevant, as far as making restitution is concerned. For the same reason, it may be noted, characteristic C will become again morally irrelevant the moment all the requirements of compensatory justice with respect to the treatment of C-persons have been fulfilled. Thus justified reverse discrimination is limited in its scope, being restricted to the righting of specific wrongs within a given range of application. Once the balance of justice with regard to C-persons has been restored, they are to be treated like anyone else. The appropriate test for the restoration of the balance of justice (that is, fulfilment of the requirements of compensatory justice) is determined by the set of criteria for just compensation that would be mutually acceptable to all who understand the unjust practice and who view the matter disinterestedly (following Rawls, as before).

### III

Does the foregoing view entail that, in the given society, *each individual* who is not a C-person has a duty to make reparation to *every* C-person he happens to be able to benefit in some way? This would seem to be unfair, since the individual who is claimed to have such a duty might not himself have intentionally or knowingly participated in the discriminatory social practice, and might even have done what he could to oppose it. It also seems unfair to C-persons, who would then be compensated only under the contingency that particular non-C individuals happen to be in a position of being able to benefit them. We must here face the questions: To whom is owed the compensatory treatment, that is, who has the right to reparation? And upon whom does the obligation corresponding to that right fall?

If we consider such compensatory policies as Affirmative Action and the Equal Opportunity Program to be appropriate ways of restoring the balance of justice, a possible answer to our questions becomes apparent. For such programmes are, within the framework of democratic institutions, social policies carried out by organized agencies of a central government representing the whole people. They are not directed toward any 'assignable' individual (to use Bentham's apt phrase), but rather are directed toward any member of an 'assignable' group (the class of C-persons) who wishes to take advantage of, or to qualify for, the compensatory benefits offered to the group as a whole. The obligation to offer such benefits to the group as a whole is an obligation that falls on society in general, not on any particular person. For it is the society in general that, through its established social practice, brought upon itself the obligation.

To bring out the moral significance of this, consider the case in which an individual has himself treated a particular C-person unjustly. By so acting, the individual in question has brought upon himself a special obligation which he owes to that particular C-person. This obligation is above and beyond the duty he has—along with everyone else—to support and comply with the social policy of reverse discrimination being carried out by his government. For everyone in the society (if it is just) contributes his fair share to the total cost of that policy, whether or not he has, personally, done an injustice to a C-person. So the individual who commits such an injustice himself has a special duty, and his victim has a special right, in contrast to the general duty of everyone to do his share (by obeying laws, paying taxes, etc.) in supporting the policy of reverse discrimination directed toward the class of C-persons as a whole, and in contrast to the general right on the part of any C-person to benefit from such a policy if he wishes to take advantage of its provisions.

The issue of the justifiability of reverse discrimination does not have to do with an individual's making up for his own acts of injustice done to this or that person. It has to do with righting the wrongs committed as an integral part of an organized social practice whose very essence was discrimination against C-persons as such. In this sense the perpetrator of the original injustice was the whole society (other than the class of C-persons). The victim was the class of C-persons as a group, since they were the *collective target* of an institutionalized practice of unjust treatment. It is for this reason that Cowan's concluding remarks do not stand up. At the end of his article he makes the following statements regarding the present members of a group which has been discriminated against in the past:

They should most certainly be treated like anyone else. But this does not mean "their" individual histories should be ignored. As with anyone else, injustices done "them" as individuals should be prevented or rectified insofar as possible. But past or future discrimination and injustice done "them" as a group and special advantages to them as a group are both out of the question since in the moral context there is no such group. (Pp. 11-12.)

But there is such a group. It is the group that was, as it were, *created* by the original unjust practice. To deny the existence of the group is to deny a social reality—a reality which cannot morally be ignored as long as the wrongs that created it are not corrected.

Cowan's position assumes that compensatory justice applies to the relations of one individual to another, but not to organized social practices and whole classes of persons with respect to whom the goals and methods of the practices are identified and pursued. This assumption, however, completely disregards what, morally speaking, is the

most hideous aspect of the injustices of human history: those carried out systematically and directed toward whole groups of men and woman *as groups*.

My conclusion is that society is morally at fault if it ignores the group which it has discriminated against. Even if it provides for compensation to each member of the group, not *qua* member of the group but *qua* person who has been unjustly treated (for whatever reason), it is leaving justice undone. For it is denying the specific obligation it owes to, and the specific right it has created in, the group as such. This obligation and this right follow from the society's past use of a certain characteristic or set of characteristics as the criterion for identification of the group, membership in which was taken as a ground for unjust treatment. Whatever duties of justice are owed by individuals to other individuals, institutionalized injustice demands institutionalized compensation.

*Brooklyn College, City University of New York*

## REPARATIONS TO WRONGED GROUPS

By MICHAEL D. BAYLES

**I**F a group of people (blacks, women) has been wronged by its members' being discriminated against on the basis of a morally irrelevant characteristic, is it morally permissible to use that characteristic as a basis for providing special considerations or benefits as reparations? It is frequently argued that since the characteristic is morally irrelevant, its use as a basis for providing reparations must also constitute wrongful discrimination.

James W. Nickel contends (ANALYSIS 32. 4, pp. 113-14) that such reparations are not wrong because they are not based on the morally irrelevant characteristic. Being black, for example, is a morally irrelevant characteristic for discriminating against or for a person. However, Nickel claims that if a black man receives special consideration as reparation for past discrimination the basis is not the morally irrelevant characteristic of his being black. Instead, it is the morally relevant one of his having been 'subject to unfair treatment because he was black' (p. 114).

J. L. Cowan criticizes Nickel (ANALYSIS 33. 1, pp. 10-12) for even including the morally irrelevant characteristic as part of the complex predicate on the basis of which reparations are given. It is not a man's having been subject to unfair treatment because he was black which is the basis of reparation, Cowan contends, 'but simply the fact that, and

the extent to which, he was unjustly discriminated against for whatever reason' (p. 11). We would not, Cowan points out, wish to favour a person who has been discriminated against for being black over another who has been discriminated against for being Jewish. He further claims that the problem of using a morally irrelevant characteristic as the basis for reparation

arises when rather than individuals it is the group which is intended, and individuals are regarded merely as members of that group rather than in their individuality. This creates a contradiction since the original premise of the moral irrelevance of blackness on the basis of which the original attribution of unjust discrimination rests implies that there is and can be no morally relevant group which could have suffered or to which retribution could now be made. (P. 11.)

The solution of the problem proposed by Nickel and Cowan, that the reparation is based on a characteristic other than the morally irrelevant one, is spurious. By parallel reasoning it can be argued that the original discrimination was not on the basis of a morally irrelevant characteristic. Racists do not discriminate against blacks simply because they are black. Rather, they claim that blacks as a class are inferior in certain relevant respects, e.g., they lack certain abilities and virtues such as industriousness, reliability and cleanliness. Thus, reasoning similarly to Nickel and Cowan, racists could contend that they do not discriminate on the basis of a morally irrelevant characteristic, but the morally relevant ones which are thought to be associated with being black. Further, a reformed racist could contend that he was mistaken to believe blacks lacked such abilities and virtues. But, since it was a *mistaken* belief which was the basis of his discrimination, he was not responsible and owes no reparations. In short, if being black is not the basis for reparations, it was not the basis for the original discrimination.

However, there need be no contradiction involved in claiming that being black is both morally irrelevant for discriminating against people and morally relevant in discriminating in favour of people to provide reparations. One may simply hold that there is no justifiable moral rule which, when correctly applied, supports discriminating against blacks, but there is one which supports discriminating in favour of them. One may hold that people have an obligation to give reparations to groups they have wronged. By using the characteristic of being black as an identifying characteristic to discriminate against people, a person has wronged the group, blacks. He thus has an obligation to make reparations to the group. Since the obligation is to the group, no specific individual has a right to reparation. However, since the group is not an organized one like a state, church, or corporation, the only way to provide reparations to the group is to provide them to members of the group.

Being black can, thus, become morally relevant in distinguishing between those individuals who are members of the group to whom reparations are owed and those who are not. But being black is only derivatively morally relevant. It is not mentioned as a morally relevant characteristic in the rule requiring one to provide reparations to groups one has wronged. Instead, it becomes relevant by being the identifying characteristic of the group wronged. The way in which it is an identifying characteristic here differs from the way in which it is an identifying characteristic for the racist. Being black is an identifying characteristic for the racist only because he thinks it is contingently connected with other characteristics. But being black is not contingently connected with the group one has wronged. Rather, it is logically connected as the defining characteristic of members of the group.

Cowan has failed to distinguish the relevance of being black as applied to groups and individuals. One may hold that one owes reparations to the group, blacks, not because the group is the group of blacks, but because the group has been wronged. But with respect to individuals, one may, as reparation, discriminate in their favour on the basis of their being black. One discriminates in favour of individuals because they are black, but one owes reparations to the group, blacks, because it has been wronged. Nor does such a position commit one to favouring individual blacks over individuals who belong to another wronged group, e.g., women. One has a morally relevant reason to favour each on the basis of their being black and female respectively. It does not follow that either should be preferred over the other.

Nothing has been said to support accepting any moral rules or principles applying to groups. (As a matter of fact, most people appear to accept one such rule, namely, that genocide is wrong.) Nor has anything been said to support accepting the rule that one has an obligation to provide reparations to groups one has wronged. These remarks have only indicated how being black can derivatively be a morally relevant characteristic for discriminating in favour of individuals if such a rule is accepted.

*University of Kentucky*



## INDIVIDUALS, GROUPS AND INVERSE DISCRIMINATION

By ROGER A. SHINER

MANY morally sensitive people find themselves faced with the following dilemma. On the one hand, they are persuaded by the argument that if being black, e.g., is morally irrelevant, then it is morally irrelevant and no more justifies favourable inverse discrimination than it justifies unfavourable discrimination. On the other hand, this move seems to open the way to neglect, whether benign or malign, of genuine social injustices. James W. Nickel (ANALYSIS 32.4, 113-4) and J. L. Cowan (ANALYSIS 33.1, 10-12) have done much to bring the logic of this situation to the surface. I shall not resist their general strategy of showing that the above is a false dilemma. My concern is with Cowan's diagnosis of the trouble as consisting in the illegitimacy of the thought that blacks *as a group* deserve inverse discrimination. His view is that one cannot argue that blacks as a group deserve retribution without also implying blackness as such is a morally relevant characteristic. This is false.

In the first place, it simply is not true that sense can never be made of the thought that a group as such deserves inverse discrimination. Consider these cases. (1) The Illyrians, through the incompetence of their negotiators, entered the European Economic Community under extremely unfair conditions. Later on, an E.E.C. Council member argues, 'We ought now to give the Illyrians especially favourable consideration', and recommends inverse discrimination. (2) Form 3B are not allowed to go on the school outing because there are 35 in the class—or so they are told, and this is not a matter of the school bus size or the tickets available. The form master argues later, '3B should be given special consideration', and recommends inverse discrimination.

The objection might be raised that Form 3B is not really a group in a sense relevant to the problem at issue. It is an individual member of the group 'classes in elementary schools', and an individual school-class may deserve inverse discrimination because of unfair treatment in the same way as an individual black or woman. However, there are many educators who argue that a disproportionate amount of money has been spent in recent years on secondary and post-secondary education, and that we now need to spend an equally disproportionate amount of money on elementary schools and schoolchildren as a group. But my argument need not turn on the accuracy of this remark, for the Illyrians are not a closed formally defined group as is Form 3B, nor are they members of some wider group.

Now it might be said that statements about Illyria or the Illyrian

nation are reducible without remainder to statements about individual Illyrians. So I still have not presented a counter-example to the thesis that a group as such cannot deserve inverse discrimination. But this is a highly contentious philosophical thesis, and the debate about whether it is true remains unresolved. It cannot be that the present question about inverse discrimination is simply this old chestnut in a new guise, and indeed Cowan does not talk as though this is what he has in mind. Those who wish to support inverse discrimination draw on the stock of available collective nouns, and frequently speak of the Black/Indian/Jewish nation/race/people and of the female sex.

These points show, then, I submit, that talk about groups deserving inverse discrimination, as opposed to similar talk about individuals, cannot be simply ruled out as nonsensical, nor will it do to rule it in purely on the basis of some reductionist theory about nations. If we are to get at what is peculiar about the thought that blacks as a group deserve inverse discrimination, we must try a different tack.

Consider these remarks—(3) 'My car ought to go in for repair, because it is a 1970 Ford'; (4) 'George deserves inverse discrimination, because he is black'. Cowan and Nickel are upset about (4), because it is logically of a piece with (5) 'George deserves to be discriminated against, because he is black', and we want to reject that inference, on the grounds that the feature mentioned is irrelevant. But why should my car's being a 1970 Ford and just exactly that mean that it ought to go in for repair? We explain a case like (3) by treating it as an elliptical argument, with a missing premiss (3\*) to the effect that all 1970 Fords have been recalled by the maker and ought to go in for repair. We can then underpin (4) in the same way, by supplying a premiss (4\*) to the effect that all blacks have been discriminated against unjustly in the past and deserve discrimination now.

This is pretty clearly the kind of point Cowan wants to make, but to show that (4) won't do unless underpinned with (4\*) is not to show what if anything is wrong with (4) 'Blacks as a group deserve inverse discrimination', still less that it is the introduction of the notion of a group deserving inverse discrimination that is problematic. As I have implied, to get from the need for (4\*) to the illegitimacy of (4), we will at least need to grapple with the reductionist position outlined above.

Nonetheless, the need for (4\*) will enable us to get a correct picture of the peculiarity of the bald inference (4). Compare (3)/(3\*) and (4)/(4\*) with (6) 'This figure is a rectangle, because it has equal diagonals'. (6) to the same degree seems to need (6\*) 'All figures with equal diagonals are rectangles'. However, although the (3) and (4) sets and the (6) set are in this respect structurally similar, there is an important difference in their content. (3\*) and (4\*) are, if true, then *a posteriori* true. (6\*), on the other hand is true and moreover *a priori* true. Thus, the arguments

constituted by the (3) set and the (4) set will only be sound as well as valid, if (3\*) and (4\*) are as a matter of empirical fact true. The difference between the (6) set and the (3) and (4) sets is that there is an *a priori* and conceptual connexion between being a rectangle and having equal diagonals, whereas there is *no a priori* and conceptual connexion between being a 1970 Ford and needing to go in for repair, nor between being a black and deserving inverse discrimination.

*This*, then, I take to be the fundamental point about the moral irrelevancy of the characteristic of being a black. It is not a matter of whether it is a group or an individual that is held to have the characteristic. It is instead a matter of the kind of link that exists between the possession of that characteristic and the possession of some other moral characteristic. The moral irrelevancy of being a black is a matter of the absence of the appropriate *a priori* link.<sup>1</sup> Contrast (7) 'Fred deserves inverse discrimination, because he is socio-economically disadvantaged', and the corresponding (7\*) 'The socio-economically disadvantaged deserve inverse discrimination'. In this case, many philosophers would be prepared to concede that, if (7\*) were part of a plausible moral theory, e.g., *à la* Rawls, (7\*) would state a conceptual connexion.

In short, my thesis in this paper is that, if we want to get clear why being a black is morally irrelevant though blacks deserve inverse discrimination, then the individual/group distinction is a red herring. We need instead the *a posteriori* connexion/*a priori* connexion distinction. The absence of the latter connexion means that any claim to the effect that blacks, whether as a group or as individuals, deserve inverse discrimination must stand, not simply on their being blacks, but on the facts of history. If, that is to say, blacks as a group deserve inverse discrimination, it will be because some claim like (4\*) is true. But if (4\*) is true, then one can reasonably argue that blacks as a group deserve retribution *without* also implying blackness as such is a morally relevant characteristic. I suspect the facts of history can stand the weight thus put on them.

<sup>1</sup> The idea that there are *a priori* connexions between moral and non-moral features is not a new one, though it is still regarded by many as heretical. I have said some of the things I would want to say by way of defence of this notion, as far as concerns connexions of the general form 'P is a reason for Q', in *Dialogue* X (1971), 759-67.

## OUTCOMES AND ABILITIES

By BERNARD GERT and JAMES A. MARTIN

IN our article ‘“What a Man Does, He Can Do”?’ (ANALYSIS 33.5) we argued that the commonly accepted view that if a man does perform a feat this shows that he has the ability to perform that feat is mistaken. A man may perform a feat by luck, in spite of his lacking the ability to do that thing. To say that a man has the ability to do a given thing, as opposed to saying that he did it, or will do it, is to say that given a reasonable opportunity he will, if he wants, do it a reasonable number of times. If the act in question is difficult and demanding, his doing it once in the right sort of circumstances may show that he has the ability; if the act is simple and easily done by luck, one success usually will do nothing to show ability. Further, in no case is ability *entailed* by a single success on the basis of some principle of modal logic that relates ‘does’ and ‘can’, as Geach had maintained in *Mental Acts*.

Geach attempts to defend his view in his note appended to our paper. But his success is doubtful since it is by no means clear that his talk about what is ‘within a man’s range’, what a man ‘can do’ and ‘could do’, is relevant to questions about abilities. Fortunately, his note is short enough (four sentences) for us to be able to reply to it in some detail and thus bring out more clearly the kinds of mistakes involved.

(1) Professor Geach says ‘If a man does perform a feat, that shows conclusively that he can perform it, that the feat is within his range; it does not merely show that his having performed the feat is logically possible (and it is quite gratuitous on the authors’ part to conjecture that I may have been confusing capacities with logical possibilities).’ Geach is certainly right that if a man performs a feat, that shows more than that his performing it was logically possible; it also shows that it was physically possible, and if this is what Geach means by ‘within his range’, it also shows that it was within his range. But if S’s doing X once (even if done by a fluke) ‘shows conclusively’ that X is within S’s range, then X’s being within S’s range has little or nothing to do with S’s having the ability to do X. It may have been gratuitous of us to say that Geach confused logical possibility with ability, for as we admit above it may have been physical possibility that he confused with ability. But we doubt if any possibility other than logical possibility is clear in modal logic, let alone ‘clear if anything in modal logic is clear’, and this remark misled us. But in any case it seems that some such confusion must be presumed to account for Geach’s attempt to use modal logic to justify going from ‘S does X’ to ‘S has the ability to do X’.

(2) ‘If a man has once done a thing, he *could* do it, however much this

may surprise him and others, and whether or not he can do the like again if he tries; anyone who bet beforehand that the man couldn't do the thing has lost his money.' This seems to be Geach's argument, and is the crux of the dispute. If we bet that someone does not have the ability to say correctly what card is on the bottom of the deck, have we lost the bet if he gets it right once? Geach says yes, we say no. Suppose that Smith offers to bet Geach that he can tell him what card is on the bottom and Geach takes the bet. Suppose further that Smith gets it right the first time, but wrong the next fifty times. Has Geach won or lost? Or suppose that Smith is wrong the first time, but right the next fifty. Has Geach won or lost? It depends on the bet. If Geach had bet that Smith *would* not get the card right, then he has lost the first, but won the second. But if he had bet that Smith did not have the ability to tell the bottom card, he would have won the first and lost the second. Notice that even if the bet concerns only the first card, the decision as to whether the man had the *ability* to tell what the first card was is to be made in the light of such things as his overall record of success and failure. We think it would be generally agreed that he has the ability in the case where he is *wrong* the first time but right the next fifty, and that he lacks it in the case where his only success comes on the first try. (Cf. Austin in 'Ifs and Cans', *Philosophical Papers*, p. 166n.) But if the bet concerns only Smith's getting the first card right and not his ability, then the winning and losing situations are reversed. Of course, when we bet, we almost always bet on outcomes and not abilities. This is why it is plausible to say that the man who bets 'Smith can't' loses if Smith does even once. But suppose that we want a man to do magic tricks; a man, say, who can tell us the bottom card of the deck. Now we focus on questions about abilities. Will we want to hire the man who has this trick within his range, in the sense that is guaranteed by a single success? Imagine a booking agent arguing to collect his fee for a magician who succeeded once out of fifty tries: 'I promised to provide a man who could do it; and that he *could* do it was conclusively shown by his single success—however much this may have surprised him and others.' A bit of a nasty surprise to the man who has to pay! In such cases, questions as to what things are within a man's range are of no interest; what we want to know about when we hire is *ability*.

All of this goes to show that deciding philosophical questions on the basis of whether we would say 'he can', or 'see, he could', is a very tricky business. And for an obvious reason: these locutions can mean quite different things. If I say 'I told you he could' did I mean that he *would*, or did I mean that he had the ability? If we are not careful to specify in advance whether we are betting on what a man *will* do, or on what he has the *ability* to do, we may get taken by someone who does not overlook these distinctions.

(3) Geach's next remark is puzzling. "There is of course an important distinction between what a man is shown able to do when he does it by fluke, or again under extreme pressure like a threat of execution, and what a man can regularly do.' In order to maintain the contrast that Geach seems to want, we presume that by 'what a man can regularly do' he meant 'what a man is shown able to do when he does it regularly'. What is shown when a man does X regularly is that he has the ability to do X. It is less clear what Geach thinks is shown when a man does X by fluke. It may be that Geach holds that doing X by fluke (e.g., telling what the bottom card is) shows that the doer has the ability to do something other than X, (e.g., the ability to utter the card names) and, as luck would have it, X happens (the right card turns up). But it is obvious that showing that a man has the ability to utter the card names does nothing to show that he has the ability to tell what the bottom card is. On the other hand, perhaps Geach wants to say that the lucky success at doing X shows that the doer has the "ability to do X by fluke", i.e., a "fluke-ability" to do X. But it is far from obvious that this has any clear sense (to put it mildly). We suspect that the fluke-ability to do X is nothing more than the physical possibility of doing X. If it is, then one might hold on the basis of modal logic that doing X by fluke does entail that one has the fluke-ability to do X. (But what would it mean to say of a normal person that it is physically possible for him to tell what the bottom card is?) Briefly: it may be that doing X by fluke always shows that the doer has either the "fluke-ability" to do X, or the ability to do something other than X; but neither of these shows that he has the ability to do X.

(4) Geach properly recognizes that this has little bearing on his main point. He says 'But this is doubtfully relevant to my discussion in *Mental Acts* of concepts as mental capacities; it has (to put it mildly) no clear sense to say of a man "It was only by a fluke, or under grievous threat, that he performed that judgment—he had not the *concepts* that would normally be needed even to think of the matter".' The reason why this has no clear sense is given quite explicitly in our paper. We suggest that for *mental* acts, we do not count something as a performance of that act unless we are willing to attribute to the putative agent the ability to perform that act. (Cf. Wittgenstein on reading, *Philosophical Investigations*, I, §157). This, as we say, may be a special feature of the concept of a mental act. And, of course, if the genuineness of a single performance rests upon ability, then not too surprisingly it makes no clear sense to say the feat was genuine but the person did not have the required ability.

As we pointed out in our paper, we have no quarrel with Geach's view of concepts as mental abilities; some confusion may have arisen because our primary interest was not in Geach's conclusion. What we

objected to is his going from 'does' to ability on the basis of the logic of the modal word 'can'. Geach's reply is worth detailed response because in defence of the old mistake new errors emerge which are interesting in their own right, e.g., failing to distinguish between 'would do' and *ability* in the betting case, and confounding the ability to do X with either the "fluke-ability" to do X or the ability to do Y (which leads to X by luck). Both the original mistake and the new ones are well worth correcting.

*Dartmouth College and University of Wyoming*

*Professor Geach writes:*

Fortunately my reply to Mr. Gert and Mr. Martin can again be brief. My numbering corresponds to theirs.

(1) Of course 'A can bring it about that *p*' is not to be confused with 'It is physically possible that *p*' any more than with 'It is logically possible that *p*'. I did not commit either confusion: what I said, and still say, is that from 'A brings it about (say, at time *t*) that *p*' there follows 'A is able, it is within A's range of accomplishments, (at time *t*) to bring it about that *p*'.

The inference is drawn that if I apply the rule of modal logic that *esse* implies *posse*, then 'some such confusion must be presumed' on my part as the confusion between ability and logical possibility. This inference must be unwarranted unless modal logic of possibility can deal only with *logical* possibility. There exists a similar view that the logic of the existential quantifier applies only to *logical* existence; even the gentle Neville Keynes called this a pernicious error; to my mind the same applies to the authors' view, and they give no reason why I should change my mind.

(2) 'A can bring it about that *p*' may mean (i) 'A can *right now* bring it about that *p*' or (ii) may imply some sort of *generalization* about when A can bring it about that *p*. The authors seem to be aware of the distinction; but if they imply that I was unaware of it, they are mistaken; I now say only that they *seem* to be aware of it, for they wrongly identify 'A *can* right now . . .' (e.g. in a betting situation) with 'A *will* right now. . .'. If A refuses to try, he has not made good his claim that he can, but neither has the claim been refuted. It is doubtful if a distinction has really been grasped when one side of it is badly misstated.

(3) What I meant by 'what a man can regularly do' is: 'what it is true of a man as a regular thing that he can do'; not (as the authors say) 'what a man is shown able to do when he does it regularly'. Conversing in Chinese may be something a man *can* regularly do, but not something he regularly *does*, because most of the time there is no fellow speaker of

Chinese around. The authors' misconstruction makes it unnecessary to discuss their further remarks under this head.

(4) I am glad the authors admit my *ab esse ad posse* inference to be valid for mental performances. They have given not a shadow of reason for denying that it holds good for *all* performances.

*University of Leeds*

## JUSTICE IN A TRIAL

By MICHAEL LEAHY

PROFESSOR Anscombe and Mr. Feldman (AF) obviously feel that the Report on Evidence of the Criminal Law Revision Committee is better met by ridicule than by serious argument (ANALYSIS 33.2). Following the heavy irony of the 'some might say' in what is, presumably, their own position:

Perhaps some might say that the justice of a trial involves something deeper than procedures designed to procure, and continue to procure, accurate results (p. 35)

they then present us with a market of clichés, heavy with the very obscurity which the report is attempting to dispel, which they assert 'define what is a just trial'.

But the problem that the committee seems to have been set is one that cannot be done away with by definition. It is an important task to look closely at those criteria which we accept, perhaps as a result of intellectual inertia, as necessary conditions of a just trial. To dismiss the implications of such attempts as 'mild torture' or 'infamous methods' is merely to lard one's inertia with indignation.

The AF view is summed up in their own words: '... "court of justice" is not the same thing as "committee for nailing criminals"' (p. 35). This has a good liberal ring to it; but it is not compelling as argument. That there is an important distinction is maintained, on the AF view, first, in a trivial way by the use of a term like 'nailing', and, secondly, by the blatant conflation of legality and morality. I will deal with each of these in a little more detail.

If we drop the silly terminology of 'nailing' and 'clobber', with its overtones of the mindless and even sadistic, which prejudice the issue, and talk, say, of 'committee for establishing legal guilt', the distinction between that and 'court of justice' will be obscured; in my view rightly. If legal guilt is established then the court has done its job, as a *legal* institution. The Report on Evidence is perhaps suggesting that we ought



to revise the implications of the revelation that in a particular case guilt was established after some impolite persuasion behind locked doors at a police station. This ought not to defeat the description of the process as a *legal* trial, nor its convening as the *legal* institution of a court of justice, unless it can be shown that as a general rule such methods produce information which is seriously unreliable. Nonetheless such revelations might well incline some to the justifiable view that the process was unpleasant, or even immoral; but not *ipso facto* illegal. If the report is urging a possible distinction of this sort then it seems a courageous, if provocative, move. It certainly has impeccable historico-legal precedents.

I am already clearly involved in my second accusation: *AF*'s unquestioning conflation of legality and morality. Although it is a commonplace that the two spheres share conceptual ground, the implications of emphasising the clear differences between the two concepts are often overlooked. The result is that the law is criticized for failing to be an outward sign of inward moral conviction before ascertaining if that is what it *ought* to be. Law reform based upon a critical examination of assumptions about the unity of justice, or the claim that a legal process which offends moral susceptibilities is bad *law*, or that a court of justice should establish moral guilt and that it is this which serves as the basis for legal liability, is, if it gets off the ground, a fascinating prospect. But even as it stands the Report on Evidence raises genuinely difficult problems of philosophical significance.

*University of Kent at Canterbury*

## ON UNDETECTABLE DIFFERENCES IN SENSATIONS

*By* CARROLL LEWIS

IT is well known that there can be *undetectable* differences in such things as colours, sounds, sizes and shapes—differences which can be known through inference to exist, but which cannot be observed directly. For example, there may be no detectable difference in colour between patches A and B, and similarly between B and C, when there is a detectable difference between A and C. Presumably there is in such a case an undetectable difference between A and B, and between B and C. Can there be undetectable differences in such things as flavours and odours (as distinct from tastes and smells)? It seems so, since, for example, it seems possible that a subject could always detect a difference in flavour between a sample of wine A and one of C, but not between one of A and one of B, nor between one of B and one of C.

Can there be undetectable differences in *sensations*, such as pains? According to the "classical" view shared by most philosophers, a sensation is so closely bound up with the sensing or feeling of it that it cannot have any aspects not themselves sensed or felt. But consider the following episode, which took place in a certain possible world.

A certain chief torturer ordered his assistant to give their uncooperative victim a pain of greater intensity than the last. The assistant had just given the victim an electrical shock by pressing the lowest-voltage button A, which was known invariably to give a subject an A-level pain. The assistant decided to press the medium-voltage button B, known to give a subject a B-level pain, which subjects were unable to distinguish from A-level pains. Knowing that saving electricity looked good on his record, the assistant avoided pressing the highest-voltage button C, which was known to give a subject a C-level pain, distinguishable from an A-level pain but not from a B-level one. For he reasoned that B-level pains must be between A-level pains and C-level pains in intensity. When the chief accused the assistant of disobeying the order, the latter changed his mind and pressed button C, giving the victim a C-level pain. After the next answer from the victim the chief ordered the assistant to give the somewhat-less-uncooperative victim a pain of the same intensity as the last; whereupon the assistant, with his concern about power consumption, pressed button B, giving the victim a B-level pain. After the next answer the chief issued another order to give the victim a pain of the same intensity as the last, whereupon the assistant pressed button A, giving the victim an A-level pain. The chief, noticing what had happened, accused the assistant of having disobeyed orders. But this time the assistant defended himself, contending not only that he had followed the orders but that his ingenious plan, if consistently pursued, would allow him continually to lower the voltages, with a consequent reduction in power usage of such proportions that he should be promoted to chief torturer.

(This hapless assistant, despite protracted torturing, maintained to the end that he was right, thus carrying with him to Hades the "classical" view of sensations to which he had become converted. Thereafter the commonsensical chief, unaware of the incoherence of his own view, was wont to cite the example of this unfortunate assistant as a lesson to all aspiring torturers who might be tempted to devote thought to the nature of sensations.)

*Southern Illinois University at Edwardsville*

## UNDERDETERMINATION OF THEORY AND INDETERMINACY OF TRANSLATION

By R. KIRK

ACCORDING to Professor Quine, two rival systems of translation could both conform not only to whatever empirical evidence actually turned up, but to all the *possible* evidence, yet still provide quite different translations of a given sentence (*Word and Object*, chapter 2; see also *Words and Objections*, D. Davidson and J. Hintikka, eds.). If this thesis of the indeterminacy of translation is true, and is not misinterpreted so as to appear merely trivial, many of our ordinary assumptions about meaning, synonymy, analyticity, and indeed about a whole cluster of other notions (including, e.g., that of belief) are radically mistaken. So far Quine's thesis seems to have resisted all efforts to disprove it.<sup>1</sup> Yet the arguments advanced in its support have been far from conclusive. In view of the interest and importance of the thesis, therefore, a fresh defence from Quine himself ('On the Reasons for Indeterminacy of Translation', *Journal of Philosophy*, 67 (1970), 178-183) demands careful examination.

Strictly, what Quine argues for in his recent paper is not the general thesis of translational indeterminacy, but the following:

- (T) Translation of physical theories is indeterminate at least to the extent that physical theories are underdetermined by all the possible data.

Suppose we are setting out to translate a foreigner's language and physical theory from scratch—"radical" translation. Starting from the assumption that two different physical theories could both be compatible with all the possible data, Quine contends that (where *A* and *B* are such theories) it might be possible for us to 'adopt *A* for ourselves and still remain free to translate the foreigner either as believing *A* or as believing *B*' (180). Nothing could possibly show that the foreigner did in fact believe one of the two theories rather than the other. Indeed, 'the question whether . . . the foreigner *really* believes *A* or believes rather *B* is a question whose very significance I would put in doubt' (180*i*). (Notice how the implications of his thesis ramify beyond the limits of the philosophy of language.) As he says, if you see physics as underdetermined only in its highest theoretical reaches, 'then by the argument

<sup>1</sup> Including one by the present writer (*Mind*, 78 (1969), 321-341). See A. Hyslop, 'Kirk on Quine on Bilingualism', and R. Harris, 'Translation into Martian', both in *Mind*, 81 (1972). D. Goldstick and others, in correspondence, pointed out my error soon after publication.

at hand I can claim your concurrence in the indeterminacy of translation only of highly theoretical physics' (181). But his own view is that 'the empirical slack in physics extends to ordinary traits of ordinary bodies'; and from this by no means implausible premiss his present argument, if sound, yields the remarkable thesis of indeterminacy of translation in its full generality.

My aim is to show that Quine's argument for the restricted thesis (T) fails, and therefore provides no support for his general indeterminacy thesis.

Quine's main assumption is that physical theories are at least to some extent underdetermined even by all possible evidence. If we consider all the observation sentences of the language, he says, and apply dates and positions to them in all combinations, without regard to whether observers were at the place and time, some will be true and some false simply on account of 'the observable though unobserved past and future events in the world'. And 'physical theory is underdetermined even by all these truths' (179). He paraphrases the point thus:

Physical theories can be at odds with each other and yet compatible with all possible data even in the broadest sense. In a word, they can be logically incompatible and empirically equivalent. (Loc. cit.)

One might object that these sentences provide not a paraphrase but a further point. It is one thing to say that two different theories might be empirically equivalent; another, surely, to say that two different but empirically equivalent theories might be logically incompatible. Why should some pairs of theories not be so diverse that there was no sense in saying that sentences of one member of a pair actually conflicted with sentences of the other? However, we need not pursue the question: Quine's argument requires only the assumption that *different* theories, not necessarily incompatible ones, could be empirically equivalent. For if rival systems of translation ascribe different theories to the foreigners, those systems must be logically incompatible—provided the theories are different in some non-trivial sense. The proviso is essential, but it raises a difficulty for Quine. For there is an obvious but unacceptable sense in which it might conceivably be said that two empirically equivalent theories were different, even logically incompatible. Theory *A* might be compatible with all possible data, and theory *B* might be identical with *A* except for differences in theoretical vocabulary. *B* might be derivable from *A*, for example, by simple interchange of the words (phoneme-sequences) 'electron' and 'neutron' wherever they occurred. Then some sentences of *B* would be negations of sentences of *A*. But in this example *B* is not a different theory from *A* except in an unacceptably trivial sense: there could be no genuine choice between attributing *A* or *B* to a foreigner. If (T) and the general indeterminacy thesis are to be

philosophically interesting, therefore, mere differences of vocabulary (or mere phonetic differences) must not be allowed to count as theoretical differences. (See also below, p. 199). Quine does not discuss suitable conditions for theoretical identity in the article under discussion, nor, so far as I know, anywhere else. The required sense of 'theory' would have to be one in which the same theory could be expressed in different sentences; yet he uses his indeterminacy thesis as a justification for attacking the philosophical respectability of notions (e.g., 'proposition') which seem to belong to the same family as 'theory'. It is therefore unclear whether he can elucidate a notion of theoretical identity without either trivialising both (T) and his general indeterminacy thesis, or being inconsistent. (It might be suggested that Quine need not worry if the notion of theoretical identity presupposed by his main assumption is unsatisfactory: if that assumption leads to the indeterminacy thesis and the latter to the assumption's falsity, surely the whole thing may be treated as a *reductio*? But this suggestion simply ignores the point of Quine's present argument, which is to help to show that the indeterminacy thesis—or at least a restricted version of it—is *true*. It could provide no support whatever for the indeterminacy thesis unless the assumption about different but empirically equivalent theories were intelligible and, indeed, probably true.) But now let us examine his argument itself, having noted that it seems to sit rather insecurely upon its main assumption.

If we are aiming at the radical translation of a radically foreign physicist's theory, Quine says:

The starting point is the equating of observation sentences of the two languages by an inductive equating of stimulus meanings. In order afterwards to construe the foreigner's theoretical sentences we have to project analytical hypotheses, whose ultimate justification is substantially just that the implied observation sentences match up. But now the same old empirical slack, the old indeterminacy between physical theories, recurs in second intension. Insofar as the truth of a physical theory is underdetermined by observables, the translation of the foreigner's physical theory is underdetermined by translation of his observation sentences. If our physical theory can vary though all possible observations be fixed, then our translation of his physical theory can vary though our translations of all possible observation reports on his part be fixed. (Pp. 179f.)

Now it is obvious that this argument must not require as a premiss either the general thesis of the indeterminacy of translation, or any special version of it which implies the indeterminacy of translation of physical theory. Any such premiss alone would yield the conclusion and render the argument redundant. Moreover, the argument can prove, if anything, only that the translation of physical theory is underdetermined by all possible data; it does not even purport to show that the translation

of anything else is also subject to the indeterminacy. As Quine remarks later:

What degree of indeterminacy of translation you must . . . recognize . . . , granted the force of my argument, will depend on the amount of empirical slack that you are willing to acknowledge in physics. (181. Cf. also the remark quoted on pp. 195–6 above.)

If his argument is sound, therefore, it must work even if we assume that it is only the translation of sentences above a certain theoretical level which suffers (if anything does) from Quinean indeterminacy. In other words, the argument must work if even we assume that all sentences up to a certain theoretical level (that level above which physical theory is supposed to be underdetermined) are *correctly translatable*. By describing a counter-example embodying this assumption I shall try to show how Quine's argument fails. (If I am right, the fault occurs at the point where he assumes that when we are trying to discover what theory the foreigner holds we have nothing solid to guide us but the matching up of observation sentences, so that our analytical hypotheses, on which word-by-word translation depends, are underdetermined.)

Imagine we are trying to discover what theory is held by Martian physicists. If we get one of them to instruct us in his theory, all the sentences he utters by way of non-technical introduction and explanation will be, by the above assumption, translatable correctly. So will all those sentences in which the expressions of his theoretical vocabulary are merely mentioned, as in verbal definitions. So, too, will all the non-theoretical expressions (logical connectives, etc.) of those sentences in which theoretical expressions are being used and not merely quoted or referred to. It follows that we can become as well placed as our instructor's fellow Martians when it comes to deciding what theory he is expounding.

For suppose the Martians are agreed that they all hold the same physical theory, which we call '*M*', and there is a Martian textbook expounding *M* from scratch, including all the necessary mathematics and logic. *Ex hypothesi* all the sentences in this book in which theoretical terms are not actually used are correctly translatable. Of course, comprehension of these sentences is not sufficient for acquiring a grasp of *M*: it is also necessary to grasp the theoretical sentences. But how is this possible, even for Martians? Nothing further can be required than a careful, intelligent study of the ways in which the theoretical sentences are embedded in more or less non-technical ones, and (especially) of the interconnections of the laws of the theory and the ways in which the theoretical expressions figure in these laws. Therefore we, the translators, can acquire a grasp of the Martian theory *M* without first having to translate the theoretical sentences. It is no more necessary for us to

translate the theoretical sentences before we can grasp  $M$  than it is necessary for Martian students of physics to do so.

We could make a partial translation of the Martian textbook, as follows. We put into English everything except Martian theoretical words; and (as our assumption permits) we give the appropriate English syntax to those sentences in which theoretical and topic-neutral expressions occur; but we leave the theoretical expressions in Martian. The result is a book differing from an English textbook only in that its theoretical vocabulary is unfamiliar. Now the theory  $M$  presented in this book might be isomorphic in the following sense with the English-speaking physicists' theory  $A$ : A theory presented in a foreign textbook, partially translated as described above, is in the relevant sense *isomorphic* with physical theory  $A$  if and only if the foreign and English theoretical vocabularies can be so mapped on to each other that, by word-for-word substitution according to this mapping, (i) The hitherto untranslated parts of the book yield sentences which fit into the context provided by those already translated in such a way that the result would be accepted by English-speaking physicists as a textbook of theory  $A$ ; and (ii) The same goes, *mutatis mutandis*, for whatever supplementary explanations of the theory presented in the book the foreign physicists may be *disposed* to give.

If the Martian theory  $M$  was in fact isomorphic with  $A$  in the sense explained, then the *only* difference between  $M$  and  $A$  would be one of vocabulary. But in that case  $M$  and  $A$  would be one and the same theory (except in a trivial sense). If mere difference of vocabulary constituted difference of theory, any fool could rival Einstein by substituting his own invented theoretical terms for Einstein's. Moreover, if mere difference of vocabulary were sufficient for theoretical difference, despite the sort of isomorphism described above, the conclusion of Quine's argument would be without philosophical interest or importance. (I do not know how great a divergence from this sort of isomorphism would be enough to justify the claim that  $M$  and  $A$  were different theories; but for my present purpose I need only the point that the isomorphism is sufficient for them to be identical.)

Isomorphism of textbooks, or even of all actually uttered explanations of  $M$  and  $A$ , would not of course logically ensure isomorphism of  $M$  and  $A$  in the sense defined, since dispositions might still diverge. But this normal inductive uncertainty must not be confused with Quinean indeterminacy. (See Quine's reply to Chomsky in *Words and Objections*, pp. 302-304.) By judicious questioning it would be possible for translators to become as clear about theory  $M$  as Martians are, and so to be justified in asserting its identity with  $A$ .

Quine assumes that the translator could not possibly improve his chances of discovering what theory the foreigners held 'by interrogation

in a theoretical vein, since the interrogation would take place in the foreigner's language and so could itself be interpreted according to either plan [sc. either by imputing  $A$  to him or by imputing to him rather  $B$ ] (180). He thinks that our interpretation of the explanations will already be contaminated by our analytical hypotheses, and could not even in principle be corrected. But we have seen that neither the indeterminacy thesis itself nor a special version of it covering the translation of physical theory may be included among the premisses of his argument. It follows that his present assumption would be valid only if it were also impossible for English-speaking pupils, 'by interrogation in a theoretical vein', to improve their chances of learning  $A$  rather than a different theory,  $B$ , from their instructors. If we start off in ignorance of physics, and can correctly understand the non-theoretical parts of Martian expositions of  $M$ , we are in the same position with respect to our chances of learning  $M$  rather than some different theory, as are English-speaking pupils with respect to their chances of learning  $A$  rather than some different theory. If English-speaking instructors can successfully teach their pupils  $A$  rather than  $B$ , then Martian ones can successfully teach ignorant translators their theory  $M$  rather than some other. And the former must be possible, since otherwise nobody could possibly tell whether he or anyone else held theory  $A$  rather than theory  $B$ , and it would make no sense to say that these were different theories, which in turn would make nonsense of both Quine's main assumption and his conclusion. For since his present argument is intended to establish the *truth* of (T), and depends on the assumption that there could be different but empirically equivalent physical theories, he is bound to agree that we could distinguish between different but empirically equivalent theories couched in English. Moreover, since evidence that  $M$  is isomorphic with  $A$  in the sense explained, and therefore the same theory as  $A$ , can be discovered by observation, 'interrogation in a theoretical vein'—given our assumption to the effect that there is no indeterminacy below a certain theoretical level—could help us to discover what theory the Martians hold.

I have assumed that if  $M$  is isomorphic with  $A$  in the sense explained it cannot also be isomorphic with a different theory,  $B$ . This must be correct, since if  $M$  were isomorphic with both  $A$  and  $B$ ,  $A$  itself would be isomorphic with  $B$ , and therefore (for the reasons given earlier) the same theory as  $B$ , *contra hypothesin*.

The situation I have described is not a counterexample to (T). For it leaves open the possibility that Quine is right in his belief that 'the empirical slack in physics extends to ordinary traits of ordinary bodies', in which case there could be no basis of correctly translatable non-technical language from which the foreign physical theory might be learnt in the way described. (My example is also compatible with (T)—



though only trivially—because it fails to exclude the limiting possibility that, contrary to Quine's main assumption, physical theory is fully determined by the totality of possible data: translation of physical theory might be indeterminate to *no* extent.) However, I think my example does show that Quine's main assumption provides no support for (T), and *a fortiori* no support for his general thesis of translational indeterminacy. For the case described is one where physical theory is underdetermined in the way he assumes, but this underdetermination does not give rise to any degree of indeterminacy of translation.

*University of Nottingham*

## WISHFUL THINKING AND SELF-DECEPTION

By BÉLA SZABADOS

PHILOSOPHERS who have tried to provide an analysis of the notion of self-deception have been at a loss to account for the difference between 'self-deception' and 'wishful thinking'. All of us would, no doubt, agree that there is a difference, in spite of the fact that there are some cases where, owing to insufficient knowledge of the details or circumstances, we cannot say which notion, 'self-deception' or 'wishful thinking', has a correct application. My aims in this paper are: first, to show that previous philosophical attempts to distinguish between the two notions have failed to do so; secondly, to provide a proper account of the distinction between the two notions.

### I

In the concluding remarks of his stimulating paper *Error, Faith and Self-Deception*, Patrick Gardiner suggests that

It might even be argued . . . that self-deception really comes down to no more than being mistaken with a motive: a self-deceiver is simply a man who wrongly believes something to be true which he would not have believed to be true in the absence of the particular interest in the matter concerned that he has.<sup>1</sup>

Gardiner immediately proceeds to voice a worry about this suggestion, viz., that the analysis of 'self-deception' in terms of 'being mistaken with a motive' blurs the distinction between 'wishful thinking' and 'self-deception'. The implication of what Gardiner says here is that 'wishful thinking' can also be analysed in terms of 'being mistaken with a motive'. Now, I want to argue that Gardiner's explication of 'being

<sup>1</sup> Patrick Gardiner, 'Error, Faith and Self-Deception', *Proceedings of the Aristotelian Society*, 70: 1969-1970, 221-244.

mistaken with a motive' is ambiguous. It can be taken to imply any or all of the following: (1) that a man wrongly believes something to be true because what he believes is really false, (2) that a man wrongly believes something to be true because his belief is not warranted by the evidence, (3) that a man wrongly believes something to be true because he believes it in the teeth of the evidence.

(a) On the first interpretation we have a man who believes something to be true (which is in fact false) and which he would not have believed to be true in the absence of the particular interest in the matter concerned that he has. It is true to say that this interpretation will not differentiate 'self-deception' from cases of wishful thinking where the belief held happens to be false. In this connexion it is perhaps worth remarking that ascriptions of wishful thinking are not comments about the truth value of the belief held, but rather they are comments about the lack of well-groundedness of someone's belief. From asserting correctly 'Ah, that is wishful thinking', where this is, of course, an assessment of a certain belief of some person, it does not follow that what he believes is false. Note, however, that if it is correctly said of someone that he has deceived himself into believing some proposition, it follows that the proposition in question is false.

(b) The second way of understanding the assertion that a person wrongly believes something to be true is that the person's belief is not warranted by the evidence. It is also true to say that this interpretation will not distinguish 'self-deception' from 'wishful thinking', for in both cases the belief is held without good reason. Note however that the holding of a belief without good reason is perhaps a necessary but certainly not a sufficient condition for self-deception.

(c) Interpretations (1) and (2) can be combined to express what appears to be Gardiner's view: that the self-deceiver is a man who believes something to be true which is in fact false and not warranted by the evidence, and which he would not have believed to be true in the absence of the particular interest in the matter concerned that he has. Gardiner is certainly right in thinking that this interpretation does not distinguish 'self-deception' from cases of wishful thinking where the belief held is false.

(d) It seems to me that it is interpretation (3) that it is essential to consider if we are to understand the difference between the notion of self-deception and the notion of wishful thinking. Surprisingly it is this sense of 'believing wrongly', viz., 'believing in the teeth of evidence', that Gardiner appears to have overlooked when asking the question 'What then is the distinction between self-deception and mere wishful thinking?'

In his book *Self-Deception*, Fingarette cryptically suggests that, if we wish, we may regard self-deception as a kind of wishful thinking

wherein we suppose the belief to be 'intentionally cultivated because of a wish to believe', whereas in plain wishful thinking 'the belief is not induced purposely and with knowledge that it is false'.<sup>1</sup>

There are difficulties with this attempt to distinguish 'self-deception' from 'wishful thinking'. What reason is there, we ask, for saying that the self-deceiver 'intentionally cultivates the belief' in question? The reason that Fingarette seems to give is that the self-deceiver has 'a passionate wish that something should be so' (*ibid.*). But the wishful thinker is also correctly characterized as one who believes something to be true 'because of a passionate wish that something should be so' (*ibid.*). And if having a strong wish that something should be so is a reason for characterizing the self-deceiver as intentionally cultivating his belief then surely it must also be a reason for characterizing the wishful thinker as intentionally cultivating his belief. Thus we still need an account of the difference between the two notions.

Fingarette also says that the self-deceiver's 'belief is induced with knowledge that it is false' (*ibid.*), while this is not true of the wishful thinker. But Fingarette forgets that it can also correctly be said of the self-deceiver that he does not really know that what he believes is false. Unless we insist that ignorance can also correctly be ascribed to the self-deceiver, we have blurred the distinction between someone who really knows the truth but is merely fooling others, and someone who is genuinely self-deceived. In view of this, we are still in the dark as to what the account of the difference between 'self-deception' and 'wishful thinking' comes to.

## II

To begin with, we do well to remind ourselves that 'wishful thinking' is *not* merely the expression of a wish or a hope. To say of someone that he indulges in wishful thinking is to assess some belief(s) of his from the point of view of rationality. Suppose that a young man is passionately in love with Miss X. He, naturally enough, yearns to be loved in return. She smiles at him on occasion, chats about the topics in the course they both attend and even invites him (among many others) to tea and cookies. He jumps to the conclusion that his love for her is being reciprocated. In spite of moments of gloom involving thoughts of unrequited love, he is, he tells his friend, convinced that she loves him too. This, it appears to me, is an example of wishful thinking. It is worth stressing that our young man does hold, quite genuinely, the belief that he is loved by her. He is cheerful and happy—muttering contentedly 'She loves me'. He does not say 'I wish she would return my love' or 'Would that she loved me!' or 'I hope she loves me'. These sentences are expressions of wishes and hopes and it is logically odd to

<sup>1</sup> Herbert Fingarette, *Self-Deception* (Routledge and Kegan Paul, 1969), p. 19.

wish for something that one thinks one has or hope for something that one is quite certain that one has. 'She loves me!' he asserts to his friend—expressing his *belief*.

Consider now the following additions and changes to this story. Suppose his friend, sympathetic but tough-minded, says when he hears this—'Ah, that is wishful thinking'. Suppose he continues by pointing out that the said young lady keeps frequent company with Z and they have been observed to be quite intimate. Our man is quite unshaken and tells the friend that surely this is mere friendship between Z and Miss X. They have probably been brought up together and are fond of each other. Furthermore, the said 'intimacies' were noted at parties where Miss X tends to be quite outgoing. In any case, they do not really bear upon their relationship. He persists in his belief, yet we detect signs of worry in his conduct and the occasional give-away in his conversation; sometimes he is on the verge of coming to terms with the mounting evidence but then, through an effort of will, he continues to explain it away and reassure himself. This, it seems to me, is a case which is properly called 'self-deception' and not simply wishful thinking. It is perhaps worth pointing out that our young man fell into self-deception through wishful thinking.

There are or may be points where we find it impossible to distinguish the wishful thinker from the self-deceiver. On the other hand, there are points where we can—and it is important to stress this. What are the points of resemblance between the two notions? Both the person who indulges in wishful thinking and the person who is self-deceived can be correctly described as holding a belief—for their believing something and expressing this belief are among the reasons for describing them in this way. Both hold the belief they do hold largely because they *want* to believe that their love is reciprocated. The truth of the belief would make them happy. They both have a personal stake in what they believe. In the absence of the motives that we ascribe to them, neither would believe what they presently believe. Both have motives, subjective reasons, for holding the belief that they do hold.

It might be said at this stage that surely one of the points of dissimilarity is that the man who indulges in a piece of wishful thinking does not 'rationalize', does not engage in the exercise of pseudo-rationality that is characteristic of the self-deceiver. It seems to me that this needs to be elaborated upon. Surely it is often said of someone who engages in a complicated piece of rationalization for something that he wants very much 'Ah, that is wishful thinking'. In the case of our young man—'Surely she loves me, she smiles at me often, talks to me and cheers me up when I feel depressed'. He gives reasons for thinking that he is loved by her. He jumps to conclusions from shreds of evidence motivated by his desire to be loved by her. We say 'Ah, wishful thinking' and point

out that by parity with this "reasoning" the said young lady could be said to be passionately in love with half of the campus. She is merely friendly towards him. Moreover, we say, she is well known to be in love with Z and produce evidence. It is at this point that wishful thinking is distinguished from self-deceit. The person to whom we ascribe wishful thinking does not *pervert* the procedures whereby we establish truth and falsehood. When evidence is brought before him that conflicts with his belief, he will, perhaps reluctantly, acknowledge it as counting against his belief. Thus, a crucial point of dissimilarity between wishful thinking and self-deceit is that in self-deceit the evidence is *against* the belief held. Once this is pointed out to the person involved, if he then proceeds to resist, by ingenious tactics, the natural implications of the evidence, we feel that he is self-deceived.

This account of the difference between 'wishful thinking' and 'self-deception' helps us to explain why we ordinarily speak of the self-deceiver in ways that have seemed paradoxical to some philosophers, while no philosopher seems to have thought that there is anything paradoxical about our ordinary talk about the wishful thinker. It is natural to say of our young man who deceives himself into believing that Miss X loves him, that deep down inside, he knows that he is not really loved by her. There isn't the least temptation to speak of 'knowledge' in connexion with the wishful thinker. The reason for this has to do with the fact that the self-deceiver fulfils *one* of the conditions for knowledge: he has good grounds for thinking that he is not loved by Miss X, while the wishful thinker does not have such grounds. Now, we speak of the self-deceiver's knowledge with qualification, never *sans phrase*, because certain conditions essential for full-blooded knowledge do not obtain in his case. A man who is properly said to know some proposition, takes the evidence that he has for the proposition as grounds for its truth. The self-deceiver does no such thing. Prompted by non-truth-centred considerations, the self-deceiver proceeds to explain away and reinterpret the evidence to satisfy some emotional need. That he is prompted by such non-truth-centred considerations either does not occur to him, or, if it does, he explains it away. Thus it is that lies tend to breed lies and self-deception tends to breed further self-deception.<sup>1</sup>

<sup>1</sup> Thanks are due to G. M. Greig, A. G. N. Flew and T. M. Penelhum for commenting on an earlier version of this paper.

# RUSSELL ON CLASSES AS LOGICAL FICTIONS

By STEVEN E. BOËR

## I

THE doctrine that classes are "logical fictions", which figures so prominently in Russell's early writings, relies heavily on the theory of "incomplete symbols" presented in *Principia Mathematica* (*PM*). For it is the alleged ability to eliminate class-symbols from sentential contexts in favour of talk about particulars and attributes (the basic constituents of Russell's ontology) which ultimately justifies thinking of classes as convenient fictions. And this ability is supposedly provided by the contextual definitions given in \*20 of *PM*.

Much has been written in the last few decades about the defects in Russell's treatment of one sort of incomplete symbol: namely, the definite description. But commentators have overlooked the fact that Russell's analysis of class-symbols, the other major sort of incomplete symbol, is also seriously defective. And the upshot is that Russell's claim to have shown that classes can generally be regarded as logical fictions is *not* supported by the formal apparatus of *PM* to which he appeals. In what follows I shall briefly review Russell's "no-class" theory, indicate the nature and location of its shortcoming, and examine the reasoning which led Russell astray.

## II

In discussing "higher-order" attributes—i.e., those which take as arguments other attributes—Russell recognized that certain problems are occasioned by higher-order attributes which are *not* extensional. By calling an attribute *f* 'non-extensional' (or 'intensional') Russell means that it *fails* to satisfy the condition

$$(1) \quad (\phi)(\psi) [(\phi x \equiv \psi x) \supset . f(\phi \hat{z}) \equiv f(\psi \hat{z})] .$$

If all attributes of attributes satisfied (1), hence were extensional, then there would be no obstacle to regarding a statement which attributes, say, *f* to  $\phi \hat{z}$  as being about the "extension" of  $\phi \hat{z}$ , the fictitious object  $\hat{z}(\phi \hat{x})$ . But since there are possible values of '*f*' for which (1) is false, this direct approach will not suffice.

For reasons of economy, Russell chooses to forego a separate account of intensional attributes and offers instead a device which allows one to extract, from *any* higher-order attribute whatever, a related *extensional* attribute. Given a statement of the form

$$(2) \quad f(\psi \hat{z})$$

which expresses an attribute of  $\psi\hat{z}$ , Russell constructs what he calls the *derived extensional function*: namely,

$$(3) (\exists\phi) [(x)(\phi!x \equiv \psi x) \cdot f(\phi!z)] .$$

If  $f$  is extensional, then (2) and (3) are formally equivalent. And if  $f$  is intensional, then (3) will at least be true whenever (2) is true.

Thus, by dealing always with (3) instead of the original (2), we may effectively confine our attention to purely extensional attributes of  $\psi\hat{z}$ . The way is now open to treat the ascription of any attribute  $f$  to the "class" determined by  $\psi\hat{z}$  as being in reality the ascription of the derived extensional function to  $\psi\hat{z}$  itself. And so we arrive at Russell's contextual definition of class-symbols:

$$(4) f\{\hat{z}(\psi\hat{z})\} =_{\text{Df}} (\exists\phi) [(x)(\phi!x \equiv \psi x) \cdot f(\phi!z)] ,$$

where the extensionality or intensionality of  $f$  is allegedly irrelevant.

### III

It is curious that Russell should have thought the move from (2) to (3) to be of any general value in avoiding the pitfalls posed by intensional attributes of attributes. For if  $f$  in (2) is indeed an intensional attribute, then (3) is clearly guilty of quantifying into the intensional context ' $f( )$ ' in the last conjunct, in consequence of which the inference from (2) to (3) is patently invalid. This may be seen as follows: let ' $f$ ' be '— is thought of by John'; then (2) and (3) may be rewritten in English as

$$(2a) \psi\hat{z} \text{ is thought of by John.}$$

and

$$(3a) \psi\hat{z} \text{ is formally equivalent to some predicative attribute which is thought of by John.}$$

But (2a) may well be true and (3a) be false! For John may never have thought of any predicative attribute whatever. (If John had read *PM*, he might find it very difficult indeed to think of a predicative attribute, since *PM* contains two quite different and incompatible explanations of their character.)

Thus (4) fails to provide an acceptable way of eliminating class-symbols from intensional contexts. It may be the case that

$$(5) \text{ John believes } \hat{z}(\psi\hat{z}) \text{ to contain } a.$$

but it remains an open question whether it is the case that

$$(6) (\exists\phi) [(x)(\phi!x \equiv \psi x) \cdot (\text{John believes that } \phi!a)] ,$$

even if we grant that the definition

$$(7) x \in (\phi!z) =_{\text{Df}} \phi!x$$

may be employed inside (6). John's belief that  $a$  is a member of the class

of  $\psi$ 's implies nothing whatever as regards his believing that  $a$  has some predicative property—let alone one co-extensive with  $\psi\hat{z}$ .

## IV

What led Russell to overlook these problems? Some light is shed on this matter by Russell's informal presentation of the "no-class" theory in *Introduction to Mathematical Philosophy*. In that account, no mention is made of *predicative* attributes: rather we are told that the extensional function derived from (2) is merely

$$(3b) \quad (\exists\phi)[(x)(\phi x \equiv \psi x) \cdot f(\phi\hat{z})].$$

Now (3b) implies (2) only on the assumption that  $f$  satisfies (1)—i.e., that  $f$  is extensional. But (2) appears to imply (3b) straightaway, regardless of the nature of  $f$ . Indeed, the proof seems to involve only the harmless truth that an attribute is co-extensive with itself plus a single use of existential generalization. Moreover, when particular examples of intensional contexts come to mind, it is easy to gloss over the dangers of quantifying-in by attending only to the *transparent* readings of belief-contexts, wish-contexts, etc. This is made all the easier by the fact that Russell's favourite examples of intensional contexts are precisely epistemic ones, which permit of the double reading as transparent or opaque.

But although one can all-too-easily slip into regarding the passage from (2) to (3b) as legitimate, the air of plausibility evaporates as soon as one attempts to pass to (3). For the explicit derivation of (3) from either (2) or (3b) requires the Axiom of Reducibility,

$$(8) \quad (\exists\phi)[(x)(\psi x \equiv \phi \mid x)],$$

and the assumption that  $f$  satisfies (1)—which is patently question-begging! By thus begging the question at the outset through the direct incorporation of (3) into (4), Russell is spared the necessity of having to appeal to (1) in proving propositions about classes. And this freedom from (1) generates the appearance of having achieved a general solution to the problem of eliminating class-symbols from intensional contexts.

It should be clear that the difficulties discussed above do not directly affect the *mathematical* theory of classes propounded in *PM*. For in that theory none but extensional attributes are ever ascribed to classes. But the *philosophical* doctrine of classes as logical fictions runs aground on the barrier created by intensional attributes.

*The Ohio State University*



## NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, R. G. Swinburne; J. H. Benson, Alec Fisher, Andrew Harrison, R. F. Holland, A. C. MacIntyre, Bernard Mayo, A. R. White, Peter Winch. This committee is responsible for appointing and advising the editor and for the general policy of the paper.

**SUBSCRIPTIONS.** The subscription to ANALYSIS, Vol. 33, is £1.50 for individuals in Great Britain, £1.75 for individuals abroad, and £2.00 for all institutions, post free. Each volume comprises six numbers, five of 32 pages and one of 48 pages, appearing within the academic year—in October, December, January, March, April and June. Orders should be sent to Basil Blackwell, 108 Cowley Road, Oxford, or placed with any bookseller.

**CONTRIBUTIONS.** Articles submitted for publication should be addressed to Dr. C. J. F. Williams, 1 Fossefield Road, Midsomer Norton, Bath, BA3 4AS. Contributors are asked to note the following requirements.

Articles should normally not exceed 3,000 words in length. **Occasionally**, however, longer contributions can be accepted;

They must be typewritten in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

**It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:**

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgement is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope and **international reply coupons** of the requisite value, whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

The copyright of articles printed in ANALYSIS remains the property of the author, but contributors are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.